ARE ALL *MTH*-LIKE GENES INVOLVED IN LIFESPAN DETERMINATION?

ANA RITA CARVALHO ARAÚJO

Dissertação de Mestrado em Bioquímica

Universidade do Porto
Faculdade de Ciências
Instituto de Ciências Biomédicas Abel Salazar
2012

ANA RITA CARVALHO ARAÚJO

ARE ALL *MTH*-LIKE GENES INVOLVED IN LIFESPAN DETERMINATION?

Dissertação de Candidatura ao grau de Mestre em Bioquímica da Universidade do Porto

Orientador – Doutor Jorge Vieira
Categoria – Investigador
Afiliação – IBMC

2012

ACKNOWLEDGEMENTS

To my supervisor, Professor Jorge Vieira, and to Professor Cristina Vieira for receiving me in the Molecular Evolution Lab at IBMC. I am very grateful for the guidance and advice that they gave me through this long journey. Thank you for making me a real scientist.

To my tutor, Professor Pedro Alexandrino, for all the support and the availability to answer all my questions and for encouraging me to never give up.

I would like to thank to my mother Olga and my father Manuel for always supporting my decisions and be by my side no matter what happened, and to my little sister Marta for all the conversations, arguments and comprehension and for always being there for me when I needed to explode.

To my closest friends, Joana and Cátia, who always have a word to say when I lost my way, and for always being there to answer all my doubts and helping me to become a better person.

To all my Master colleagues, especially, Elísio, Pedro, Andreia and Jéssica for being patient with me and for the friendship that grew up so fast within a year.

To all my lab colleagues at IBMC, especially Ana Estefânia and Inês for all the coffee breaks, conservations and good mood that we shared.

Finally, I would like to acknowledge FCT for funding this work.

ABSTRACT

The *Drosophila methuselah* (*mth)* gene was described for the first time in 1998 as being associated to increased lifespan. A mutant line of *mth* lived 35% longer than the wild type flies at both 25ºC and 29ºC. Moreover, *mth* flies were resistant to oxidative stress and starvation.

The *mth* gene encodes a putative G protein-coupled receptor (GPCR), showing a recognizable Mth ectodomain containing ten conserved cysteine residues in the N-terminal and a seven-transmembrane (7tm) domain in the C-terminal, which are called Mth typical features. The function of this protein is not known yet, but despite of being implicated in lifespan, studies in *D. melanogaster* showed evidences that it might be involved in the control of the synaptic efficacy at glutamatergic neuromuscular junctions, enhancement of resistance to stress, higher wing beat frequency and coordinated visuomotor entrainment during simulated flight, as well as maintaining the rate of stem cell division with age.

mth gene is only present in *melanogaster* subgroup of species and it has an estimated age of 10 million years. In *D. melanogaster, mth* belongs to a family comprising 16 genes: *mth* and its paralogs, *mth*-like genes. However, this family of genes only has two members in *Bombyx mori* and seven in *Anopheles gambie*. For that reason, we here establish the history of *mth*-like family through the *Drosophila* genus and concluded that the number of members of this family is not retained between *Drosophila* species. Hence, it is of interest to determine if other *mth*-like genes can influence lifespan too in species that are distantly related to *D. melanogaster* such as *D. americana* – since association between naturally occurring Mth amino acid polymorphisms and lifespan were showed in *D. melanogaster.*

D. americana has been diverging from *D. melanogaster* for about 40 million years and does not have a *mth* orthologous gene copy. However, here we show that markers for *mth*-like genes with the typical Mth protein features are able to explain a considerable amount of lifespan differences observed in a F2 association study. Consequently, this suggests that *mth*-like genes might be involved in lifespan determination in a species distantly related to *D.*

melanogaster, and furthermore they could be associated with lifespan in all the insect orders where they were described.

RESUMO

O gene *methuselah (mth)* de *Drosophila* foi descrito pela primeira vez em 1998 como estando associado ao aumento da longevidade. Uma linha mutante deste gene vive cerca de 35% mais do que as moscas do tipo selvagem a 25ºC e a 29ºC. Além disso, as moscas *mth* são mais resistentes ao stress oxidativo e à restrição calórica.

O gene *mth* codifica um recetor acoplado a proteína G (GPCR), apresentando um domínio Mth que possui dez resíduos de cisteína conservados no seu N-terminal e um domínio 7-transmembranar no C-terminal. Estas caraterísticas são denominadas como as características típicas da proteína Mth. A função desta proteína ainda não é conhecida, mas além de estar implicada em estudos de longevidade em *D. melanogaster*, mostrou evidências de que poderá estar envolvida no controlo da eficácia das sinapses na junção neuromuscular glutamatérgica, no aumento a resistência a stress, a uma maior frequência do batimento das asas assim como na coordenação visuomotora durante o voo, bem como na manutenção da taxa de divisão das células estaminais com a idade.

Este gene só se encontra presente no subgrupo de *melanogaster* e tem uma idade estimada de cerca de dez milhões de anos. Em *D. melanogaster,* o gene *mth* pertence a uma família composta por 16 membros: o gene *mth* e os seus parálogos, os genes *mth*-like. No entanto, esta família de genes só possui dois membros em *Bombyx mori* e sete em *Anopheles gambiae.* Por isso, foi determinada a história da família *mth*-like através do género de *Drosophila* e concluído que o número de membros desta família não é mantido entre as espécies de *Drosophila*. Consequentemente, coloca-se a questão se outros genes *mth*-like também influenciam a longevidade numa espécie que está distantemente relacionada com *D.melanogaster*, como por exemplo, *D. americana*. Esta questão coloca-se porque associação entre polimorfismos aminoacídicos, que ocorrem naturalmente na proteína Mth, e a longevidade foram descritos em *D. melanogaster.*

D. americana encontra-se a divergir de *D. melanogaster* há cerca de 40 milhões de anos e não possui nenhum gene ortólogo do gene *mth*. Contudo, aqui mostra-se que marcadores para genes *mth*-like que possuem as características típicas da

proteína Mth conseguem explicar uma parte considerável das diferenças observadas na longevidade de um estudo de associação F2. Por conseguinte, isto sugere que genes *mth*-like poderão estar envolvidos na determinação da longevidade de uma espécie distantemente relacionada com *D. melanogaster.* Além disso, estes genes podem também estar associados com a longevidade em todas as ordens de Insetos onde foram descritos.

Texto escrito conforme o Acordo Ortográfico - convertido pelo Lince

RÉSUMÉ

Le gène *methuselah* (*mth*) de *Drosophila* a été décrit pour la première fois en 1998 comme étant associé à l'augmentation de la longévité. Une ligne mutante de ce gène vit 35% plus que les mouches du type sauvage à 25°C et 29°C. En outre, les mouches *mth* sont plus résistantes au stress oxydatif et à la restriction calorique.

Le gène *mth* codifie un récepteur accouplé à la protéine G (GPCR), présentant un domaine Mth qui a dix résidus de cystéine conservés dans son N-terminal et un domaine 7-transmembranaire dans le C-terminal. Ces caractéristiques sont désignées comme les caractéristiques typiques de la protéine Mth. La fonction de cette protéine n'est pas encore connue, mais, en plus d'être impliquée dans les études de longévité dans *D. melanogaster*, elle a montré des évidences comme pouvant être impliquée dans le contrôle de l'efficacité des synapses dans la jonction neuromusculaire glutamatergique, dans l'augmentation de la résistance au stress, dans une plus grande fréquence du battement des ailes, ainsi que la coordination visuo-motrice au cours du vol et la continuité du taux de la division des cellules staminales avec l'âge.

Ce gène n'est présent que dans le sous-groupe de *melanogaster* et a un âge estimé d'environ dix millions d'années. Dans *D. melanogaster*, le gène *mth* appartient à une famille composée de 16 membres: le gène *mth* et ses paralogues, les gènes *mth*-like. Toutefois, cette famille de gènes n'a que deux membres dans *Bombyx mori* et sept dans *Anopheles gambiae*. Ainsi, on détermine l'histoire de la famille *mth*-like à travers le genre de *Drosophila* et on constate que le nombre des membres de cette famille n'est pas maintenu parmi les espèces de *Drosophila*. Par conséquent, une question se pose: savoir si d'autres gènes *mth*-like influencent aussi la longévité d'une espèce lointainement liée à *D. melanogaster*, comme, par exemple, *D. americana*. Cette question se pose parce que l'association entre les polymorphismes aminoacides, qui se trouvent naturellement dans la protéine Mth, et la longévité ont été décrits dans D. *melanogaste*r.

D. americana est en train de s'éloigner de *D. melanogaster* il y a environ 40 millions d'années et n'a aucun gène orthologue du gène *mth*. Cependant, ici on

montre que les marqueurs pour les gènes *mth*-like, qui ont des caractéristiques typiques de la protéine Mth, peuvent expliquer une partie considérable des différences observées dans la longévité d'une étude d'association F2. Par conséquent, ceci suggère que les gènes *mth*-like peuvent être impliqués dans la détermination de la longévité d'une espèce lointainement liée à *D. melanogaster*. En outre, ces gènes peuvent aussi être associés à la longévité de toutes les ordres d'insectes où ils ont été décrits.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Prelude

Lifespan is a life-history component that varies significantly among organisms in natural populations (Promislow *et al*., 1996). The increased or decreased *Drosophila* lifespan has been studied when some genes are manipulated (reviewed by Paaby and Schmidt, 2009), but the major factors that contribute for a genetic variance for longevity in natural populations are still uncharacterized (Paaby and Schmidt, 2008). For instance, many induced mutations in single genes belonging to the insulin/insulin-like growth factor-like signaling (IIS) pathway result in lifespan changes in divergent species such as nematodes, *Drosophila* and mouse (Giannakou and Partridge, 2007). Nevertheless, these genes may be under strong purifying selection and thus may contribute little to the observed variation in lifespan in natural populations (Paaby and Schmidt, 2008).

In *Drosophila melanogaster*, some complementary tests using mutant strains have shown that naturally occurring variation at many genes affects lifespan (reviewed by Paaby and Schmidt, 2009), including *methuselah* (*mth*) gene. The *mth* gene was described for the first time in 1998 as being associated to increased lifespan (Lin *et al*., 1998), and for this, *mth* gene was proposed as a candidate gene that may be implicated in lifespan determination. A mutant line of *mth* lived 35% longer than the wild type flies at both 25ºC and 29ºC. Moreover, *mth* flies were resistant to oxidative stress, heat and starvation (Lin *et al*., 1998). Additionally, studies in *D. melanogaster* showed that the protein encoded by this gene may be implicated in many roles as regulation of neurotransmitters exocytosis in larval glutamatergic neuromuscular junctions (Song *et al*., 2002), higher wing beat frequency and coordinated visuomotor entrainment during simulated flight (Petrosyan *et al*., 2007), as well as, maintaining the rate of stem cell division with age (Wallenfang *et al*., 2006).

The mutant line has a P-element inserted into the third intron of the *mth* gene (Lin *et al*., 1998). Both the precise and imprecise excision of the P-element were generated in *mth* flies and caused a reversion back to the wild type phenotype. Two additional lines were isolated with a deletion of the DNA adjacent to the P-

element insertion site and revealed to be null alleles of *mth* gene. In embryos, they are homozygous lethal, which suggests that *mth* plays a major role in early development.

mth gene encodes a protein whose predicted sequence shows a leader peptide and seven hydrophobic regions, suggestive of a transmembranar domain. Then a gapped Blast search revealed that Mth protein is similar to G protein-coupled receptors (GPCR) (Lin *et al*., 1998), a large and important group of receptors proteins involved in signal transduction (Schmidt *et al*., 2000). In *Drosophila*, GPCRs are classified into four groups: Rhodopsin-like receptor family, Secretin-like receptor family, Metabotropic glutamate receptor-like family and Atypical family (Brody and Cravchik, 2000). The gene *mth* is the founder member of *methuselah*-like (*mth*-like) family, a subfamily inside the *secretin*-like receptor family, which is found in *Drosophila* but not in *Caenorhabditis elegans* or vertebrates. Accordingly to Nordström *et al*. (2009), the *mth*-like family in *D. melanogaster* comprises fifteen members including the *mth* original and fourteen paralogous called *mth*-like genes. Little is known about the other members of the *mth*-like family and all the studies about *mth* gene were performed in *D. melanogaster.*

1.2. GPCRs and signaling through G proteins

GPCRs are the largest family of membrane proteins that mediate most of the cellular response to hormones and neurotransmitters and are responsible for many senses like vision, olfaction and taste. Their classical tridimensional structure comprises a seven spanning α- helical segments separated by alternating intra and extracellular loops (figure 1). Each GPCR has unique signal-transduction activities that involve various kinds of G-proteins in order to establish communication between the internal and external environments of the cell. After the binding of an agonist, the GPCR is active and can activate a specific heterodimeric G protein leading to the transmission of the signal through the modulation of downstream effectors. Every ligand triggers a specific signaling pathway via stabilization of an exclusive collection of molecules, allowing the correct interaction with them and which then will defuse a biological response (reviewed by Rosenbaum *et al*., 2009).

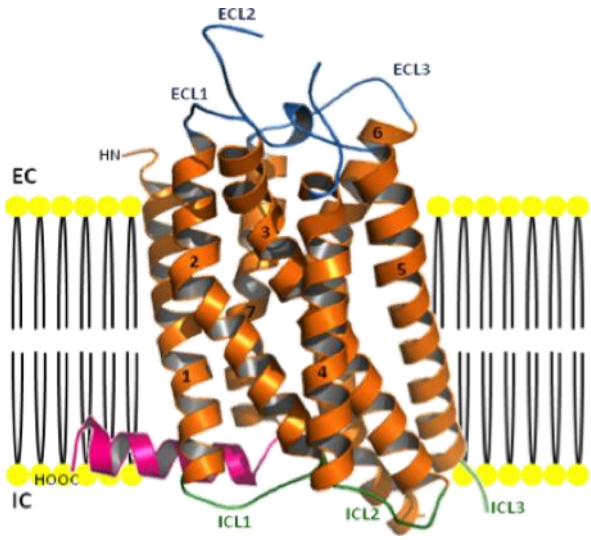

Figure 1: Representation of crystal structure of the human GPCR A2A receptor. The protein crystal was obtained by Jaakola *et al.* in 2008 and this image was created in PyMol (Schrodinger, 2010). In orange are marked the seven transmembrane domains with respective numeration, in green the three intracellular loops (ICL) and in blue the three extracellular loops (ECL). In magenta is represented the eighth helix of human A_{2A} receptor, which is in a parallel position relatively to the cellular membrane and is the intracellular region of the protein. Protein is represented in cartoon.

Like all GPCRs, the members of Mth-like family are coupled to intracellular GTP-binding proteins (G proteins) in order to transmit the signal through the cell. The G proteins are composed by three subunits: Gα, Gβ and Gγ. Once bound to ligand, the GPCR suffers a conformational change and acts as a guanine exchange factor (GEF) inducing the swap of GDP for GTP. Once the GTP binds Gα, the G proteins dissociate into Gα and Gβγ components, modulating an intracellular signal by activation of downstream effectors and second messengers in the cell (reviewed by McCudden *et al.*, 2005, Strader *et al.*, 1994). Gα and Gβγ subunits can interact with different downstream effectors, activating different signaling pathways (figure 2).

There are three ways to inactivate the signal transducing by G proteins: GTP hydrolization, through regulators of G protein signaling (RGS) proteins or by G protein coupled receptor kinases (GRKS). In the first form, the intrinsic guanosine triphosphatase (GTPase) activity of Gα subunit causes the hydrolysis of GTP to GDP, leading to its inactive state. The reassociation of subunit Gα and dimmer Gβγ terminates all effectors interactions. The second form is through RGS proteins, which are able to act like a GTPase accelerating protein (GAP) in order to facilitate the hydrolyzing of GTP and therefore deactivate Gα subunit (Neubig,

2002). In the third form, GRKs phosphorylates the active GPCR, converting it into a target for high affinity binding Arrestin followed by Clatherin-induced endocytosis of the receptor, blocking the activation of G-proteins (Wilden, 1995, Krupnick *et al*., 1997).

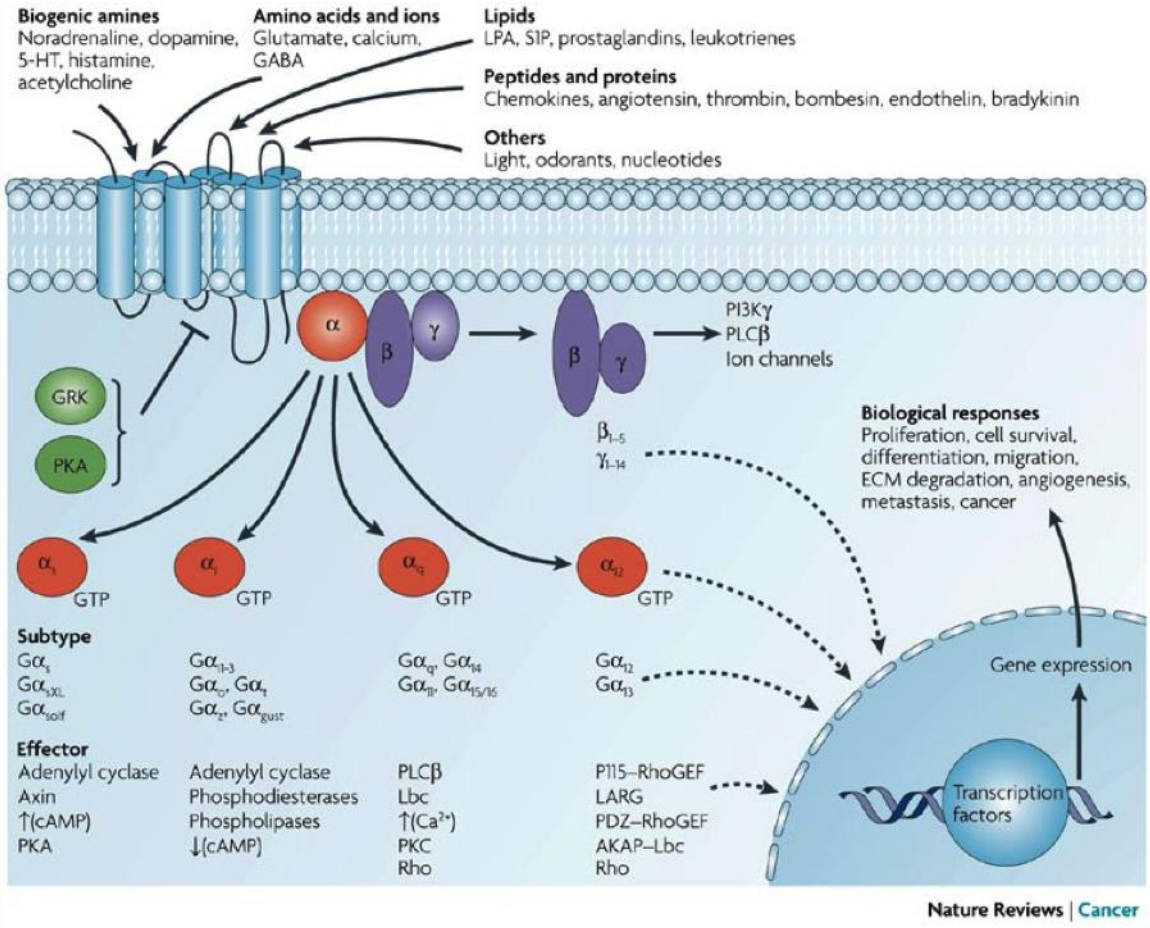

Figure 2: Diversity of G protein-coupled receptor signaling (Dorsam and Gutkind, 2007).

GPCRs are membrane proteins and are highly expressed once they are responsible for mediating the interaction between cells and their environment. Signals such as light, odors and taste or intracellular stimuli as hormones, peptides and neurotransmissors are detected by GPCRs. Since so many types of ligands activate them, they play a role in many biological processes like neurotransmission, growth, development, cellular differentiation, inflammation and the immune response (Horn *et al*., 1998). For example, in *Anopheles gambiae,* GPCRs mediate the many sensory pathways that are related with malaria transmission, because the olfactory sense plays an important role in the choice of

human host (Hill *et al*., 2002); and in humans the over expression of GPCRs and their activation by agonists released by tumor or stromal cells is the tactic most used by tumor cells to stimulate GPCRs and the signaling pathways activated by them (Dorsam and Gutkind, 2007).

Due to the fact that these membrane receptors control key physiological functions, their dysfunction contributes to some of the most prevalent human diseases and GPCRs are the target, both directly or indirectly, of about 50 to 60% of all therapeutic agents nowadays (Pierce *et al*., 2002).

However, when comparing invertebrates with vertebrates, the latter ones have a larger number of GPCRs than invertebrates, and thus vertebrates may be more sensible to sensory information from the environment that allows for a more complex homeostasis regulation. A large amount of diversified GPCRs members might enable a better adaptation of the individual to new environments and new situations (Strotmann *et al*., 2011).

1.3. The Secretin-like Receptor Family of GPCRs

The GPCRs are one of the largest protein families in mammalian genomes with about 800 members in human genome (reviewed by Lagerstrom and Schioth, 2008) divided into five groups in GRAFS classification: Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2 and Secretin (table 1) (Fredriksson *et al*., 2003). These five families are present in most of bilateral species including *Caenorhabditis elegans*, *Drosophila melanogaster, Anopheles gambiae, Strongylocentrotus purpuratus, Ciona intestinalis* and the vertebrate species (Fredriksson and Schiöth, 2005, Whittaker *et al*., 2006, Schiöth *et al*., 2007).

The *Drosophila* genome contains approximately 200 genes coding for GPCRs (Adams *et al*., 2000) classified into four families based on primary and secondary structure predictions: Rhodopsin-like, Secretin-like, Metabotropic glutamate-like and Atypical seven-transmembrane proteins (table 1) (Brody and Cravchik, 2000). The Secretin-like family includes many hormone receptors such as receptors for secretin, calcitonin, vasoactive intestinal peptide, parathyroid hormone and related peptides. These receptors are characterized by long NH_2-terminal domains

with five cysteine residues highly conserved that form disulfide bonds and by a short cytoplasmatic domain (Brody and Cravchik, 2000).

Table 1: Summary of mammalian and *Drosophila* GPCRs reviewed by Fredriksson *et al*. 2003 and Broady and Cravchik 2000. BAIs, brain-specific angiogenesis-inhibitory receptors; CALCR, calcitonin receptor; CASR, calcium-sensing receptor; CELSRs; cadherin EGF laminin-A seven-pass G-type receptors; CRHRs, corticotropin-releasing hormone receptors; EMRs, EGF-like module containing, GCGR, glucagon receptor; GHRHR, growth hormone-releasing hormone receptor; GIPR, gastric inhibitory polypeptide receptor; GLPRs, glucagon-like peptide receptors; GRM, metabotropic glutamate receptors; LECs, lectomedin receptors; PACAP, pituitary adenylyl cyclase-activating protein; PTHR, parathyroid hormone receptors; SCTR, secretin receptor; TAS1 and TAS2, taste receptors; VIRP, vasoactive intestinal peptide receptor.

Mammalian GPCRs	Subgroups
Glutamate	GRM, GABA receptors, CASR, TAS1
Rhodopsin	α, β, γ and δ
Adhesion	CELSRs, BAIs, LECs and EMRs
Frizzled/Taste2	Frizzled receptors, TAS2 receptors
Secretin	CALCR, CRHRs, GCGR, GIPR, GLPRs, GHRHR, PTHR, PACAP, SCTR, VIPR

Drosophila GPCRs	Subgroups
Glutamate-like receptors	Glutamate receptors, GABA receptors, Calcium receptors
Rhodopsin-like receptors	Opsins, GPCRs for biogenic amines, related compounds and purines, peptide GPCRs
Adhesion (Atypical)	Stan and Boss
Frizzled (Atypical)	Frizzled-like proteins
Secretin-like receptors	Hormone receptors, *mth*-like receptor family latrophilin-like receptor

In *D. melanogaster* the latrophilin-like receptor gene is thought to have a role in the control of synaptic exocytosis and was identified as a member of Secretin-like receptor family (Brody and Cravchik, 2000). Its endogenous ligands are unknown but the neurotoxin α-latrotoxin binds to latrophilin-like receptor and stimulates

the massive neurotransmitter release which leads to nerve terminal degeneration (Krasnoperov *et al*., 1997, Holz and Habener, 1998). This evidence supports the hypothesis of a putative role of Secretin-like receptor family in neutransmission. In 2003, Johnson *et al.* showed that another member of Secretin-like receptor family is indeed a functional diuretic hormone 44 receptor in *Drosophila*, once the expression of this receptor in HEK cells elevate the intracellular calcium concentrations in response to diuretic hormone 44.

The Mth-like family is a subfamily within Secretin-like family and phylogenetic studies showed similarities between this subfamily and the Adhesion GPCRs (Nordström *et al*., 2009). The evolutionary link between Adhesion GPCRs and Secretin GPCRs suggests that the latter family could also play an important role in the development of nervous system and hence *mth*-like genes (Nordström *et al*., 2009).

1.4. Methuselah and Methuselah-like family

In 1998, when Lin *et al.* described the relationship between lifespan and *mth* gene, it was also predicted that the sequence of this gene encoded a leader peptide (Mth ectodomain) and seven hydrophobic regions which suggested that it had a transmembrane domain. Using Blast tool, this sequence showed high homology with GPCR (Lin *et al*., 1998). Brody and Cravchik first annotated the *mth*-like family in 2000 in *D. melanogaster*. At that time, ten *mth*-like genes were identified and a phylogenetic tree, showing their evolutionary relashionships, was created using the neighbor joining method (figure 3) (Brody and Cravchik, 2000).

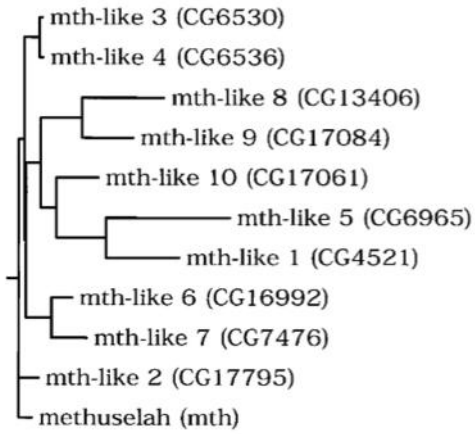

Figure 3: Phylogeny of the *mth*-like family performed by Brody and Cravchik, 2000.

In 2009, Nordström *et al.* described the other members of *mth*-like family, which from that moment till now comprises fifteen members, including *mth* gene and

fourteen *mth*-like genes (Nordström *et al*., 2009). However, details on the age and the way the *mth*-like family arose are still unavailable. This family of genes was only studied in insect species and the typical number of *mth*-like genes in insect genomes is also unknown, once *Bombyx mori* has only two *mth*-like genes (Fan *et al*., 2010) and *Anopheles gambiae* has seven (Hill *et al*., 2002), in contrast with the fifteen members present in this family in *D. melanogaster*.

In 2001, the crystal structure of Mth ectodomain was resolved by West *et al*., as well as the organization of Mth protein: an NH_2-terminal ectodomain followed by seven-pass transmembrane domains. In addition, it was discovered that the NH_2-terminal domains of the Mth-like proteins shared between 27% and 65% of sequence identity with Mth ectodomain, suggestive of a similar folding into tertiary structures (West *et al*., 2001). The Mth ectodmain has ten conserved cysteine residues that form five disulfide bonds in all Mth-like proteins, with the exception of Mth-like 3.

In 2003, Duvernell *et al.* demonstrated that *mth* ranks among the fastest evolving genes in *Drosophila* genome by examining patterns of polymorphism and divergence at this gene in samples of *D. melanogaster, D. simulans* and *D. yakuba.* This study inferred too that *mth* and *mth*-like genes are structurally and functionally related and may be involved in similar transduction pathways (Duvernell *et al*., 2003).

1.5. Methuselah and Lifespan

After being dubbed as the first candidate gene for aging in *D. melanogaster* the search for *mth* function began. In 2004, Cjevic *et al.* were able to isolate two peptides that activate Mth in HEK293 cells, which correspond to two splice variants A and B of the *Drosophila* gene *stunted (sun)* product, designated as Sun A and Sun B respectively (Cvejic *et al*., 2004). The *sun* gene encodes a protein similar to the ε-subunit of F_1F_0-ATP synthases in other organisms from yeast to humans and in order to verify if Sun is a real ligand of Mth, Sun A and Sun B polypeptides were synthesized and their effect tested on Mth-dependent intracellular calcium response. The result showed that both Sun peptides were able to activate Mth receptor, but not Mth-like 1, 2, 3, 5, 8 and 9 proteins, thus

we can conclude that Sun peptides are specific endogenous ligands of Mth. The ligands for other Mth-like proteins remain unknown till now (Cvejic *et al*., 2004). Cjevic *et al.* also investigated the effect of mutations in *sun* in longevity and heterozygous mutant flies showed an increase in lifespan.

However, more recently, were described two additional peptides that are agonists of Mth: the *Drosophila* Sex Peptide (SP) and Serendipitous Peptide Activator of Mth (SPAM). But, these new agonists share low sequence homology with Sun, indicating that the activation of Mth is promiscuous (Ja *et al*., 2009). Although, the real function and the pathway where Mth receptor is involved are still uncharacterized, recent studies of other promiscuous GPCRs likely to be involved in nutrient sensing and immune response (Migeotte *et al*., 2006, Wellendorph *et al*., 2005), suggests that Mth functions as a peptide sensor *in vivo* and may elucidate the family's role in aging (Ja *et al*., 2009).

It was also proposed that Mth controls the synaptic efficiency at glutamatergic neuromuscular junctions of *Drosophila*. This hypothesis is based on the evidence that the decreasing of Mth function leads to a decrease in calcium release, reduction in synaptic size and decrease in transmission at the synapse (Song *et al*., 2002). This can cause stress resistance in motor neuron, which is critical for aging. It is known that when over expressed in motor neurons, Superoxide Dismutase extends lifespan in *Drosophila* (Parkes *et al*., 1998), so Mth might regulate the stress resistance that is passively reflected on lifespan variation.

Furthermore, Mth was implicated in the enhancement of some sensorial abilities like wing beat frequency and visuomotor reactions to moving stimuli during fly's life. Though, *mth* gene appears to affect the toughness of some behaviors more significantly than others, the reason for these differences remains unclear (Petrosyan *et al*., 2007), but may be implied in lifespan determination.

It was also showed that *mth* hypomorphic mutation exerts relatively modest, counterbalancing positive or negative effects on lifespan as well as in fertility that are temperature-dependent. This mutation seems to have identified a locus whose allelic variation has an indirect effect on the rate of aging, since *mth* extends lifespan of flies indirectly as a consequence of decreased reproductive output (Mockett and Sohal, 2006).

1.6. The *Drosophila* genus

Members of *Drosophila* genus are among the most important model species that are used in modern biology. This is due to many reasons, such as its genome size that is near 5% of the size of mammalian genome although they share many gene families, signaling pathways, tissue and organ systems (reviewed by Matthews *et al*., 2005).

Drosophila are cheap to maintain, easy to manipulate and there are many molecular tools that are very well optimized in order to perform genetic studies. Furthermore, *D. melanogaster* became the most preferable species of this genus and today is used as a model organism for the study of fundamental problems in biology (Ashburner and Bergman, 2005). This is due mostly because of the sequencing and annotation of its genome between the end of twentieth century and the beginning of twenty-first century (Adams *et al*., 2000). In 2004 was released the complete and annotated genome of *Drosophila pseudoobscura* (Richards *et al*., 2005) followed by the genome of other ten *Drosophila* species that are now publicaly available on FlyBase (http://flybase.org) (figure 4).

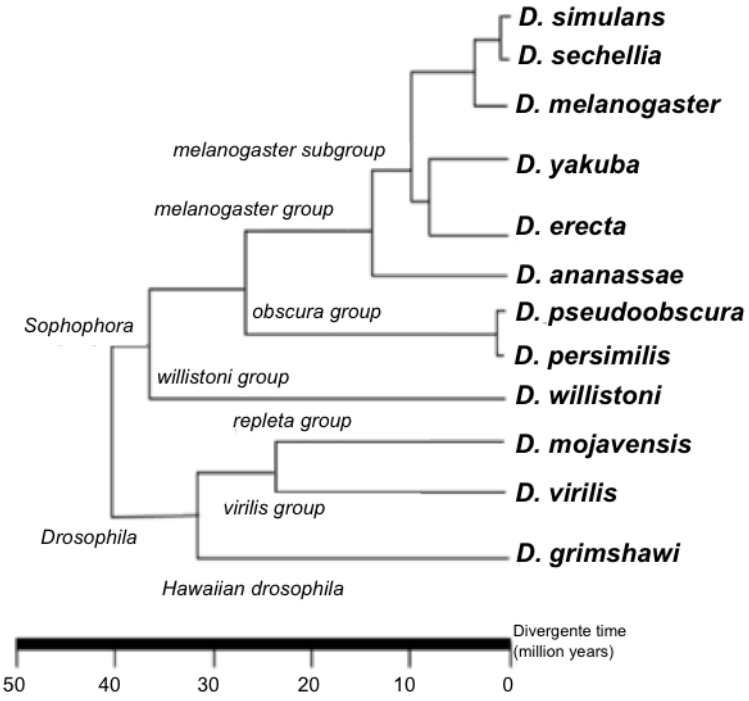

Figure 4: Phylogeny of the subset of *Drosophila* species with completely sequenced and annotated genomes

Once the *Drosophila* genus comprises two subgenera and more than 1300 species (Wheeler, 1981), it is very diverse. With this high diversity, this genus is a powerful tool to study the evolution of genes and their function in distantly related species that have a common ancestor.

1.7. *Drosophila americana* as a model organism

The genus *Drosophila* can be divided into two subgenera: *Sophophora* and *Drosophila* (figure 4). The species *D. melanogaster* belongs to subgenus *Sophophora*, while *D. americana* is part of subgenus *Drosophila.*

D. americana is a member of *virilis* group, which comprises fourteen species that are divided in four subphyla: *virilis*, *montana*, *kanekoi* and *littoralis* (Spicer, 1992, Spicer and Bell, 2002). The subphyla of *virilis* includes five taxa closely related: *D. virilis*, *D. lummei*, *D. novamexicana*, *D. a. americana* and *D. a. texana* (Throckmorton, 1982) (figure 5). Throckmorton proposed this division in 1982, based on differences such as centromeric fusions and chromosomal paracentric inversions. The complex of americana constituted by *D. a. americana*, *D. a. texana* and *D. novamexicana* is part of the *virilis* phylad.

D. americana has been diverging from *D. melanogaster* for about 40 million years (Russo *et al.*, 1995). This species is native from United States and shows a high effective population size and low levels of population structure (Schafer *et al.*, 2006, Morales-Hojas *et al.*, 2008). *D. americana* can be found in the central and eastern regions of the United States from the south (Texas to the states around the Gulf of Mexico) to the north of the country (from Montana to Maine) (Patterson and Stone, 1952).

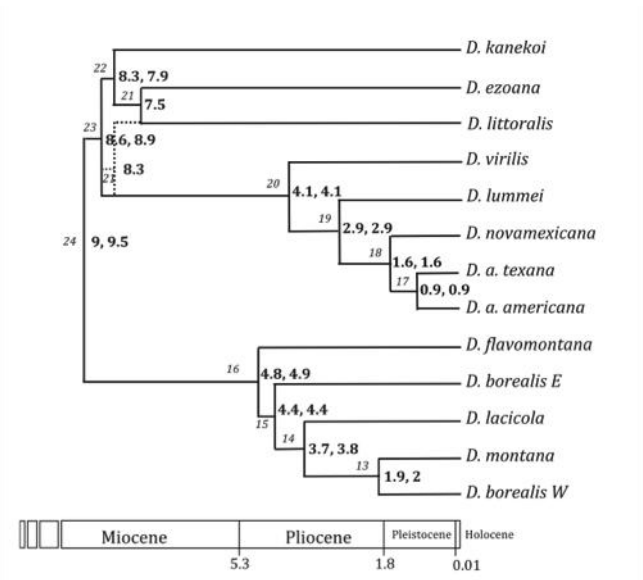

Figure 5: Chronogram showing the phylogeny of *virilis* group (Morales-Hojas et al., 2011).

This species has a centromeric fusion between chromosomes *2* and *3* and a derived *X/4* fusion chromosomal polymorphism (Hughes, 1939, Throckmorton, 1982). This latter fusion between chromosomes *X* and *4* is distributed through a very wide cline along a latitudinal gradient being at high frequency in the north of United States (*D. a. americana*) and is rare in the south of United States (*D. a. texana*) (Vieira *et*

al., 2001, McAllister, 2002). Nevertheless, several molecular studies showed that *D. a. americana* and *D. a. texana* are indistinguishable at the DNA level for genes in chromosomes *2*, *3*, *4* and *X* (Hilton and Hey, 1996, Hilton and Hey, 1997, McAllister and Charlesworth, 1999, McAllister and McVean, 2000, Vieira *et al*., 2001). However, there are some exceptions for genes *fused1*, *para, CG18543* and *Yp1* (Morales-Hojas *et al*., 2008, Vieira *et al*., 2006, Vieira *et al*., 2003, Vieira *et al*., 2001). Despite of these exceptions, it is now generally accepted that the two subspecies of *D. americana* – *D. a. americana* and *D. a. texana* – are in fact chromosomal forms and not differentiated taxa (Morales-Hojas *et al*., 2008).

It has been suggested that, in *Drosophila,* the selection regime varies according to the population latitude (Paaby and Schmidt, 2009). At high latitudes the seasonal stress imposes strong selection in flies that would be indirectly selected for a long lifespan due to stress genes with pleiotropic effects. On the other hand, in high resource availability and stable environments at low latitudes, there would be no indirect selection for long lifespan. The cline observed for *mth* haplotypes is compatible with this vision (Schmidt *et al*., 2000, Duvernell *et al*., 2003). If this is true, then the wide geographic distribution of *D. americana* means that there should be marked variation regarding lifespan in this species. Recently, a cline for a 5' flanking region polymorphism at the *regucalcin* gene was showed as being associated with abdominal size. A variant of this gene is present in high frequency in the north of the *D. americana* distribution and it is associated with small abdominal size, which is the opposite of previously described in literature that declares that large body size is common in cold places (Reis *et al*., 2011).

Hence, under the pleiotropic hypothesis for the evolution of lifespan, clines at *mth*-like genes may be expected for other *Drosophila* species apart from *D. melanogaster*. It should be taken into account that *D. americana* is polymorphic for six inversions and one *X/4* fusion with estimated frequencies higher than 5% (Hsu, 1952). The *X/4* fusion – *Xc* inversion chromosomal arrangement shows a latitudinal cline being frequent in the north of geographic distribution and almost absent in the south of the distribution, held by the weak selection (Vieira *et al*., 2001, McAllister, 2002, Vieira *et al*., 2006, Reis *et al*., 2008, McAllister *et al*., 2008, Vieira *et al*., 2003). Cold resistance was proposed to be partially responsible for the preservation of such cline (McAllister *et al*., 2008), although a cline has never been reported for the other chromosomal arrangements.

Therefore, *D. americana* seems to be proper for addressing the generality of the findings made for *D. melanogaster*, regarding the involvement of *mth*-like genes in lifespan variation.

The aim of this work is to enhance the information about this family of genes indicated as playing an important role in lifespan determination as well as in many other aspects of *Drosophila* development. For that reason, we wanted to know how this family of genes evolved through the years in *Drosophila* genus and test if these genes are involved too in the determination of lifespan in a species distantly related with *D. melanogaster – D. americana.*

2. METHODS

2.1. Identification of *mth*-like genes in *Drosophila* genus and phylogenetic analyses

In order to identify how many *mth*-like genes are present in the annotated genome of twelve species of *Drosophila* genus publically available, a tBLASTn search as implemented in FlyBase (http://flybase.org) using the amino acid sequence of *mth* gene from *D. melanogaster* as query was performed. From this search, a total of 141 sequences were obtained. Nucleotide sequence alignments were performed with the help of Clustal tool as implemented in DAMBE (Xia and Xie, 2001). The amino acid sequences were first aligned and then used as a guide to align the corresponding nucleotide sequences. Phylogenetic Bayesian analyses were run for nucleotide sequences using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). The GTR model of sequence evolution was used, thus allowing for among-site rate variation and a proportion of invariable sites. Moreover, third codon positions were allowed to have a gamma distribution shape parameter that is different from that of first and second codon positions. Two simultaneous and completely independent analyses were run for 500.000 generations (each with one cold and three heated chains), starting from random trees. Samples were taken every 100th generation. The first 1250 samples were discarded (burn-in).

Then, we rooted the resulting gene tree in order to make inferences on the age of *Drosophila mth*-like genes. This was obtained by using, in addition to the compilation of 141 retrieved protein sequences of *Drosophila,* a set of fifty proteins from non-insect species that, when using BLASTp show the highest similarity to the *D. melanogaster* Mth protein, and the Minimum Evolution algorithm, as implemented in MEGA (Kumar *et al.*, 2008). The resulting tree showed that the *D. melanogaster mth*-like 1, *mth*-like 5 and *mth*-like 14 genes are the oldest members of *mth* gene family, and therefore the Bayesian tree was rooted accordingly. Two *mth*-like sequences (one from *D. pseudoobscura* – *GA26708* – and another from *D. yakuba* – *GE26296*) were not used, once they would introduce a large number of alignment gaps and thus compromise the use of many alignment positions. It should be noted that Mth ectodomain is only

found in insects (Nordström *et al*., 2009), and for that reason, the inferred relationships are based on the non Mth-ectodomain region of the proteins.

2.2. Detection of protein domains

Protein domains were detected using the Conserved Domains tool at NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer *et al*., 2011).

2.3. F2 association experiments

Two different experiments were performed during this work. The first *D. americana* F2 association study was performed using five strains – W11, H5, O57, W29 and W46 – forming 5 crosses H5♂ x W11♀, W11♂ x W46♀, W29♂ x O57♀, O57♂ x H5♀, W46♂ x W29♀ named crosses AA, AT, TN, NA and TT, respectively. This F2 association study comprises 433 male individuals: 88 from cross AA, 75 from AT cross, 87 from TN cross, 94 from NA cross and 89 from TT cross. After the eclosion, all these individuals were collected and maintained at 25ºC under 12h light and dark cycle conditions up to their death with the aim to measure the lifespan of each individual. These 433 individuals were genotyped for all the nine *mth*-like genes present in *D. americana* genome (http://evolution.ibmc.up.pt) (see results) – *mth*-like 1, *GJ16328* ortholog, *mth*-like 11, *mth*-like 5, *mth*-like 14, *mth*-like 8, *mth*-like 9, *mth*-like 10 and *GJ12490* ortholog. The second *D. americana* F2 association study was performed using strains H5 and W11, described in detail in Reis *et al*. 2011, and was used to test for associations between *mth*-like genes and longevity too. This F2 association study is composed by 453 F2 *D. americana* males showing extreme phenotypes for developmental time, chill-coma recovery time, abdominal size and lifespan, that are descendents of three F0 H5 x W11 crosses and were selected out of 975 individuals. These 453 individuals were genotyped for 3 *mth*-like genes (*GJ12490* ortholog, mth-like 5 and *mth*-like 11). Not all individuals could be genotyped due to unknown reasons.

In the F2 association study there is only one generation of recombination and therefore associations are expected between phenotype and polymorphic sites in the region(s) where the causative variant(s) are located.

2.4. Genomic DNA extraction

Genomic DNA from single males was extracted using the QIAamp DNA Mini Kit from QIAGEN (Izasa Portugal, Lda.)

2.5. Molecular markers for *mth*-like genes

Specific primers (table 2) based on both the annotated *D. virilis* (http://flybase.org) and non-annotated *D. americana* (http://evolution.ibmc.up.pt) genomes were developed for all *mth*-like genes present in the genome of both species. For all genes, standard amplification conditions were 35 cycles of denaturation at 94ºC for 30 seconds, primer annealing (table 2) for 45 seconds and primer extension at 75ºC for 2 minutes. The DNA fragments were run in a 1,5% agarose gel stained with ethidium bromide, using SGTB buffer (Grisp, Portugal).

PCR products obtained with primers developed for *mth*-like 1*, GJ16328* ortholog, *mth*-like 5, *mth*-like 14*, mth*-like 8*, mth*-like 9 and *GJ12490* ortholog were then digested with restriction enzymes (table 2) and run in a 2% agarose gel stained with ethidium bromide, using SGTB buffer. The *mth*-like 11 and *mth*-like 10 show alleles with different sizes, and those were used as a marker.

Table 2: Molecular markers designed for all nine *mth*-like genes present in *D. americana* genome. Ta, annealing temperature; RE, restriction enzyme.

Gene	Primer Sequence (5' ➔3')		Ta (ºC)	RE
***mth*-like 1**	mthl1_GJ14760_F	CTACGACAGCGGCACCAA	61	*Msp*I
	mthl1_GJ14760_R	CCAGCAGCAGGCAGAATA		
***GJ16328* ortholog**	GJ16328_ex1_F	GTCCAGAGCACAGCAGAT	53	*Hpy*CH4IV
	GJ16328_int1_R	TAAACAAAGCGAAAAATG		
***mth*-like 5**	129ex4F	AGCCATCGTCTATTTTAT	50	*Bcl*I
	Mth_129w11poly_R	ACATCACCAGACACATCA		
***mth*-like 11**	mth_60_amer_F	TGTTTGAGGTTTGAGTTT	52	–
	mth_60_amer_R	CACGCCGAAGATGTAATA		
***mth*-like 14**	mthl4_g16090_70F	CAAAAACGGAAATGAGTC	51	*Sau*3AI
	mthl4_g16090_641R	TGGTGGTATGAAGTAACG		
***mth*-like 8**	GJ12496_mthl8_F	CTGATTGCGGCTGGCTAT	59	*Sau*3AI
	mthl8_alter_R	TGGTGAACAGATGCTCCC		
***mth*-like 9**	GJ12495_mthl9_ F	CAACAACAACTGCGACAA	57	*Sau*3AI
	GJ12495_mthl9_ R	ATTGAAGAGCCTCCGAAT		
***mth*-like 10**	GJ12492_mthl10ex1_F	TACGAACAGTTTGTGGTG	52	–
	GJ12492_mthl10ex2_R	AAGGGTATGGAGAAGAGC		
***GJ12490* ortholog**	Mth_29_F	GTTCTTTCCGAGCAGCAA	53	*Sau*3AI
	Mth_29_R	CAGAGCACACAGCAGAGC		

2.6. Cloning and DNA sequencing

The *GJ16328* gene was not localized in *D. virilis* genome, but there is a hit of this gene in *D. americana* non-annotated genome. For that reason, some experiments were performed in order to certify that it is the same gene. Firstly, flanking primers were designed based on the *D. virilis GJ16328* gene in order to amplify this gene. Standard amplification conditions were 35 cycles of denaturation at 94ºC for 30 seconds, primer annealing at 50ºC for 45 seconds, and primer extension at 75ºC for 2 minutes. An amplification product was obtained for two individuals, one from strain H5 and one from strain W11. Secondly, the PCR products were cloned using the TOPO-TA Cloning Kit for sequencing (Invitrogen,

Spain). Then, positive colonies were picked randomly, grown in 5 mL of LB with Ampicillin, and plasmids were extracted using the QIAprep Spin Miniprep Kit from QIAGEN (Izasa, Portugal). The plasmidic DNA of three colonies was sequenced in order to correct for possible nucleotide miss-incorporations that may have occurred during the PCR reaction. Sequencing reactions were performed using ABI PRISM Big Dye cycle-sequencing kit version 1.1 (Perkin Elmer, CA, USA) and the primers for the M13 forward and reverse priming sites of the pCR2.1 vector. Sequencing runs were performed by STABVIDA (Portugal).

The sequence of *mth*-like 1 was not complete in the non-annotated genome of *D. americana* because there is a gap in the sequence of this gene. So, a similar approach was used in order to have access to the full sequence. We started by designing flanking primers for *mth*-like 1 based on the sequence of this gene in *D. virilis* genome. Standard amplification conditions were those described above as well as all the process till the sequencing.

2.7. Gene expression analyses

The expression levels of seven *mth*-like genes – *mth*-like 1, *GJ16328* ortholog, *mth*-like 11, *mth*-like 8, *mth*-like 9, *mth*-like 10 and *GJ12490* ortholog – were determined for eight sets of three males each from the H5 and W11 strains at four different times of longevity – 0 days, 10 days, 30 days and 60 days. After eclosure 0, 10, 30 and 60 days old, individuals were frozen in liquid nitrogen. Total RNA was isolated from each set of individuals using TRIzol Reagent (Invitrogen, Spain) according to the manufacturer's instructions and treated with *DNase I* (*RNase*-Free) (Ambion, Portugal). cDNA was then synthesized by reverse transcription with SuperScript III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen, Spain), using random primers. No-template controls and reactions with RNA that was not reverse transcribed were performed in order to confirm the absence of DNA contamination. Specific primers with efficiency between 90% and 100% (table 3) for all the seven genes were used when performing qRT-PCR experiments using the isolated cDNA.

Table 3: Specific primers for the seven *mth*-like genes and *RpL32* tested for qRT-PCR. bp, base pair.

Gene	Primer Sequence (5' ➔3')		Fragment size
***mth*-like 1**	mth1_RT3_F	TTGAATGGAACGACGCTG	129 bp
	mth1_RT3_R	TGGCTGGAACTTGATGAA	
***GJ16328* ortholog**	16328_RT3_F	CACATTCCTTTGGCTCTT	128 bp
	16328_RT3_R	CAAGTAACCCCGCAGATA	
***mth*-like 11**	mth60_RT_F	CTCTGGTTCGTCATTTCG	139 bp
	mth60_RT_R	TATGCGTTGGTTTTCTCA	
***mth*-like 8**	mthl8_RT2_F	CTCACGCTGATACGCATC	130 bp
	mthl8_RT2_R	CGCAAACGAAATAGGTAA	
***mth*-like 9**	mthl9_RT_F	TCTCCGTGTATTTGGTGG	151 bp
	mthl9_RT_R	AGCAGCAGAATGAACCCG	
***mth*-like 10**	mthl10_RT2_F	CGGAGAGCGTCACATTGG	115 bp
	mthl10_RT2_R	CTGGCAGTGTTGGCAAAC	
***GJ12490* ortholog**	mth29_RT_F	GGAGAGAAGGAGTCGGTG	110 bp
	mth29_RT_R	GTGCGTTCATTGCTGTCG	
RpL32	RpL32_RT_F	ACAACAGAGTGCGTCGTC	208 bp
	RpL32_RT_R	ATCTCCTTGCGTTTCTTC	

All the experiments were performed in duplicate and with the iQ SYBR Green Supermix (Bio-Rad, Portugal) according to the manufacturer's instruction on a Bio-Rad iCycler with the following program: 3 minutes at 95ºC; 40 cycles of 30 seconds at 94ºC, 30 seconds at 58ºC and 30 seconds at 72ºC. Specific primers were also designed for the endogenous *ribosomal protein L32* (*RpL32*) (table 3) that was used as the reference gene. Fold change in expression was calculated using the $2^{-\Delta\Delta_{CT}}$ (Livak) method (Livak and Schmittgen, 2001).

2.8. Statistical analyses

All statistical testes and genotype-phenotype associations were tested using non-parametric tests and the software SPSS Statistics 17.0 (SPSS Inc., Chicago, Illinois).

Linear regression analyses were performed in order to estimate the percentage of variation in lifespan that can be explained by the molecular variation present at *mth*-like genes.

3. RESULTS

3.1. The rise of *mth*-like genes

3.1.1. The evolutionary origin of Mth ectodomain

After performing BLASTp searches at NCBI (http://www.ncbi.nlm.nih.gov) in non-*Drosophila* insects using Mth ectodomain from *D. melanogaster* as a query, we were able to find that Mth-like proteins with the typical Mth features (a

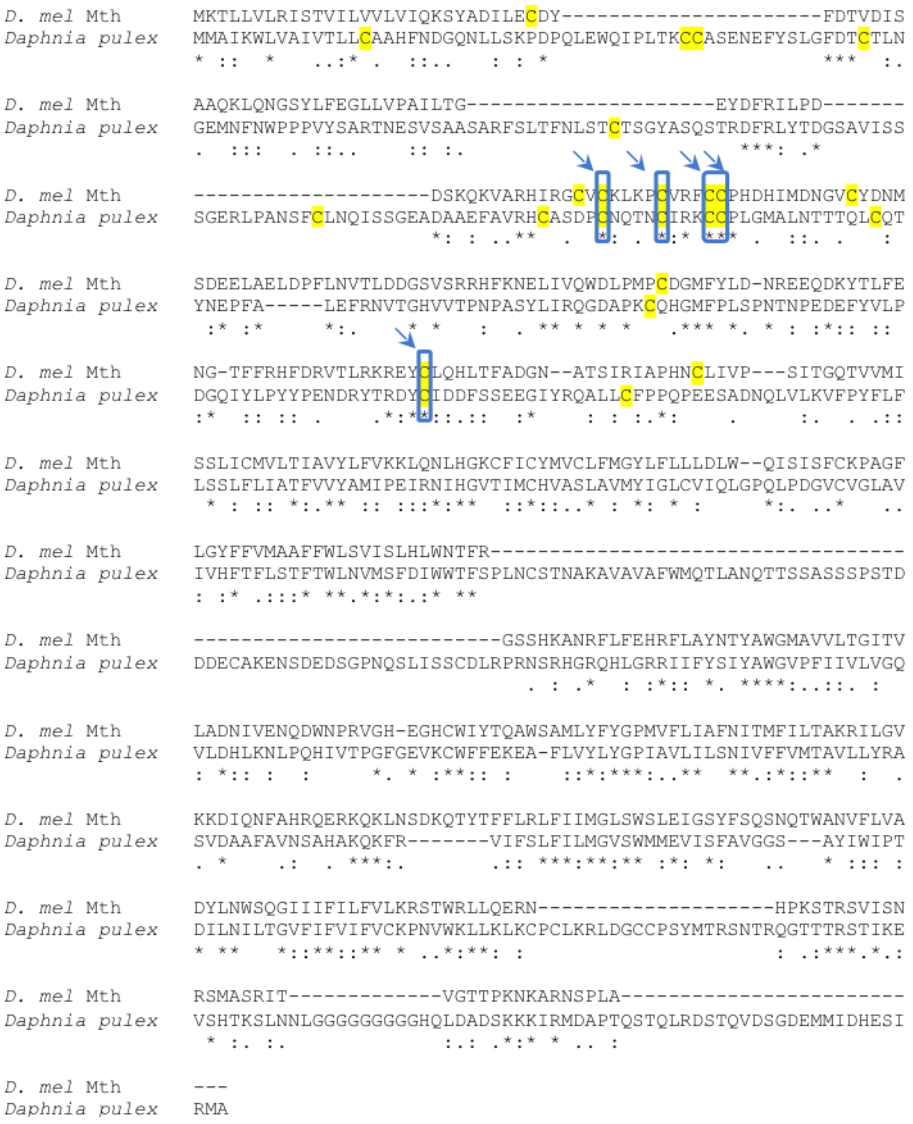

Figure 6: Sequence alignment of the Mth protein from *D. melanogaster* and a protein from *Daphnia pulex* (gi321478728). Arrows and boxes mark the five out of the ten cysteine residues of Mth ectodomain that are conserved in *Daphnia pulex.*

recognizable Mth ectodomain with the 10 typical cysteine residues, and a 7tm domain (West *et al*., 2001)) in the Diptera species *Anopheles gambiae* (gi158295564; gi15829556; gi1216692) and *Culex quinquefasciatus* (gi170057117; gi170057123), as well as in *Spodoptera frugiperda* (Lepidoptera; gi835836696), and *Acyrthosiphon pisum* (Hemiptera; gi193575614). As a result, genes encoding Mth-like proteins with all the typical features of Mth are to be found in insects that have been diverging for more than 350 million years (Gaunt and Miles, 2002, Meusemann *et al*., 2010). Using the same approach, we identified one Mth-like protein in the Crustacean *Daphnia pulex* (gi321478728). This sequence shows a 7tm domain but does not show a recognizable Mth ectodomain. Five out of the ten residues of cysteine, typical of Mth ectodomain, are however conserved. Four out the ten typical cysteine residues are not conserved, but there are cysteine residues in the close proximity (less than ten amino acid residues away) in the *Daphnia* sequence (figure 6). When a BLASTp was performed, as implemented in FlyBase (http://flybase.org), and using the *Daphnia* putative Mth-like ectodomain region sequence as a query, there are significant hits (except value greater than 0.05) with *Drosophila* proteins encoded by members of the *mth* gene family only. Therefore, the *Daphnia* sequence seems to derive from a proto *mth*-like ancestral sequence. Crustaceans, the sister group of insects, have been diverging for about 420 million years (Gaunt and Miles, 2002).

3.1.2. Identification of *Drosophila mth*-like genes

In order to identify the *mth*-like genes in the publically available and annotated genome of twelve *Drosophila* species, the FlyBase tBLASTn (http://flybase.org) option was used and the *D. melanogaster* Mth protein as a query. This approach resulted in the identification of one *D. melanogaster mth*-like gene (*CG31720; mth*-like 15) that has not been identified by Nordström *et al*. 2009, raising the number of the members of this family to 16 in this species.

The difference in the number of *mth*-like genes present in the species from *melanogaster* subgroup is very striking compared with all the other species (figure 7). Species from *melanogaster* subgroup show more *mth*-like genes (14-18) than *D. ananassae,* as well as the species from *obscura* group and *Drosophila* subgenus (5-9). It seems likely that this variation is due to some recent

duplication events, which took place after the separation of *melanogaster* subgroup, about ten million years ago.

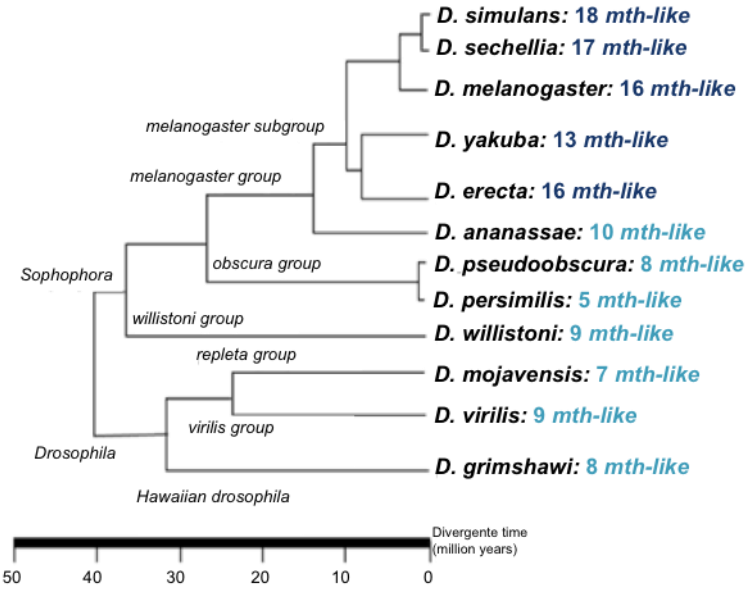

Figure 7: Phylogeny of the subset of *Drosophila* species with completely sequenced and annotated genomes and the number of *mth*-like genes present in each species.

A similar approach was used in order to identify the number of *mth*-like genes present in *D. americana* genome. A BLAST search was performed using the coding region of all *mth*-like genes present in *D. virilis* against the genome of both strains of *D. americana* – H5 and W11 – whose genome was recently sequenced. Thus, we were able to identify nine *mth*-like genes in both genomes of the two strains of *D. americana*, and as in *D. virilis* there is no *mth* ortholog in *D. americana* species.

3.1.3. Phylogenetic analyses using species of *Drosophila*

The inferred relationships, using a Bayesian approach, of the 141 gene sequences retrieved from the 12 annotated D*rosophila* genomes, are showed in figure 8. Analyses of these sequences reveal that they are all Mth-like proteins, with the exception of *GJ22009* and its orthologous in *D. mojavensis* and *D. grimshawi* that are the Diuretic hormone receptor 44 (a member of the Secretin-like family – see introduction). This observation suggests that the oldest member of the *mth*-like family is the *mth*-like 14 gene.

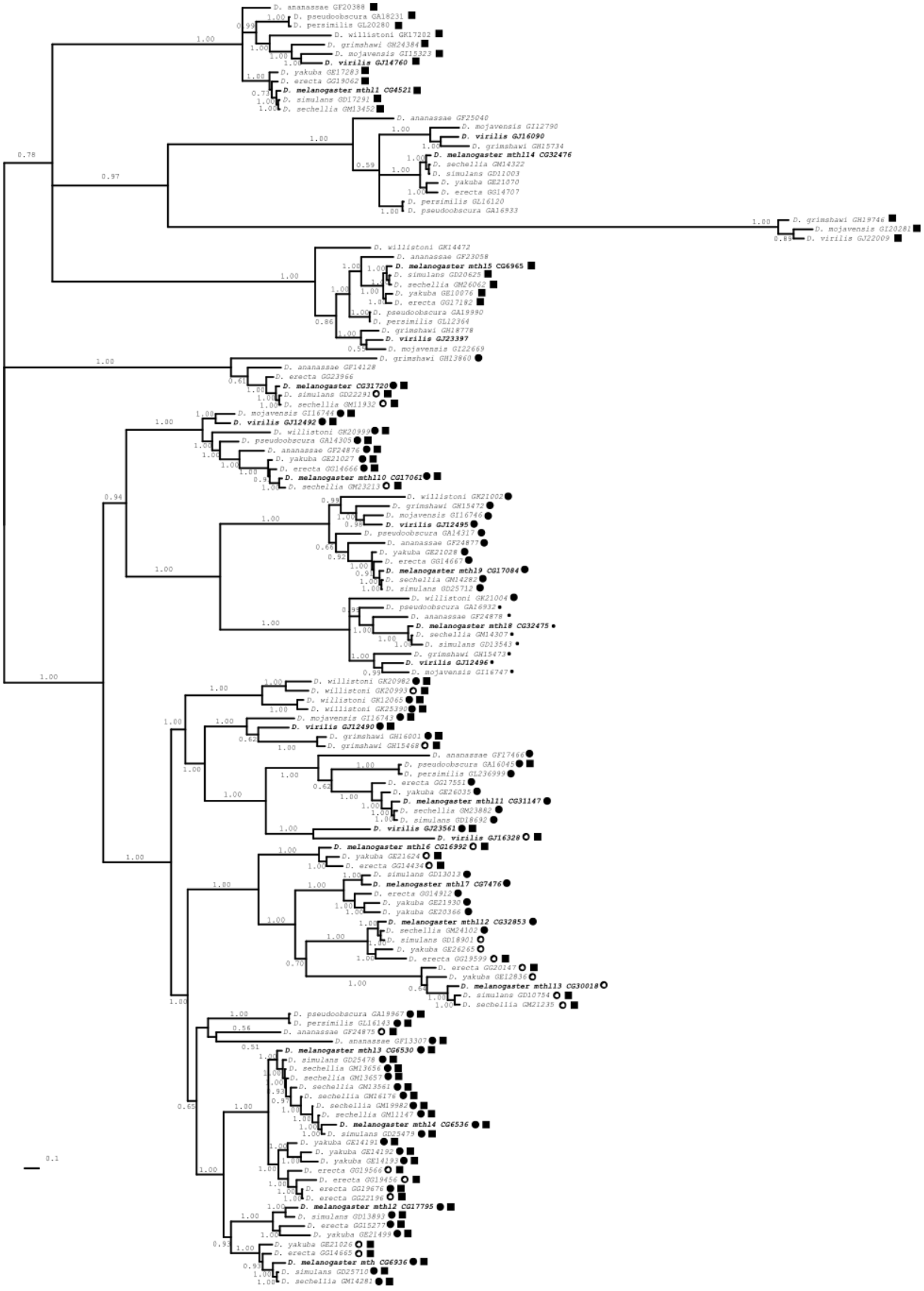

Figure 8: *mth*-like DNA sequence relationships inferred using a Bayesian approach. Values near the nodes are posterior credibility values. ●, presence of a recognizable Mth ectodomain and the typical 10 cytseine residues; •, presence of the typical 10 cysteine residues but no recognizable Mth ectodomain; **O**, presence of a recognizable Mth ectodomain without the typical 10 cysteine residues; ■, presence of a recognizable 7tm domain.

The rooted tree suggests that the Mth ectodomain evolved only once, after the separation of the *mth*-like 1, *mth*-like 5 and *mth*-like 14 gene lineages and the remaining lineages. Moreover, this domain seems to have been lost only once in the *mth*-like 8 gene lineage. It should be noted that, in the Mth-like 8 protein, the region that corresponds to the Mth ectodomain is not deleted, and that it can be aligned with that of Mth, although it seems to be divergent enough to not be identified as a Mth ectodomain. For all *Drosophila* species with the exception of *D. melanogaster*, miss-annotation of genes cannot be ruled out as a source of the apparent variation in Mth ectodomain.

The presence of a 7tm domain is an ancestral feature of the *mth* gene family, but this feature seems to have been lost several times independently during the evolution (figure 8). Once again, not being able to detect a 7tm domain does not mean that the protein is truncated. It is, thus conceivable that such proteins have a divergent 7tm domain that is not easily recognizable. Moreover, for all the *Drosophila* species, except *D. melanogaster,* miss-annotation of genes cannot be ruled out as a source of apparent variation in 7tm domain size.

After the resolution of *mth*-like phylogeny with all *mth*-like genes present in the twelve species of *Drosophila* genus, whose genomes are publically available and annotated we can see that not all *mth*-like genes have Mth ectodomain and the 7tm domain characteristic of *mth* gene.

As said before, species of the *melanogaster* subgroup show more *mth*-like genes than the other *Drosophila* species analyzed (figure 7). It is unlikely that this is an artifact of the approach used – tBLASTn using as a query the *D. melanogaster* Mth protein. Indeed, the *mth*-like genes most divergent from *mth* are the ones where orthologous gene copies could be found in all *Drosophila* species (figure 8). In *melanogaster* subgroup lineage there are many *mth*-like genes that encode proteins with the three typical features of Mth – Mth ectodomain, ten typical cysteine residues and a 7tm domain. This number for *mth*-like genes with all the typical Mth features is probably due to recent gene duplications events of such genes. *mth* is a young gene that is about 10 million years old, since it is only present in species of *melanogaster* subgroup, and the same is true for *mth*-like 2*, mth*-like 3*, mth*-like 4*, mth*-like 6*, mth*-like 7*, mth*-like 12 and *mth*-like 13. It should be noted that given the distribution of these genes in the species considered, the *D. pseudoobscura GA19967* and the *D. permisilis GL16143*

sequences are likely miss-positioned in the tree. Indeed, the support for the grouping *D. pseudoobscura GA19967, D. persimilis GL16143, D. ananassae GF24875* and *GF13307* is low (0.56).

The phylogenetic tree could be divided into three large groups: the more ancestral group where *mth*-like genes do not present Mth ectodomain; the middle-aged group where are *mth*-like genes with Mth ectodomain; and the newest group which correspond to *mth*-like genes most related to *mth* gene. None of the nine *mth*-like genes present in *D. virilis* is included in this latter group.

3.2. Characterization of *mth*-like genes

3.2.1. *mth*-like genes present in *D. virilis*

Once the model of this study is the species *D. americana*, it was necessary to annotate the *mth*-like genes present in the genome of the two strains from this species, whose genome was recently sequenced. But before that, it was important to confirm that *mth*-like genes were well annotated in *D. virilis* genome. For that, a comparison between the annotation of these genes in *D. melanogaster* and *D. virilis* was made, since the annotation in *D. melanogaster* is completely reliable, once it is based on real data (table 4).

As we can see on table 4, all *mth*-like genes present in *D. virilis* genome seem to be well annotated with the exception of *GJ16090.* In order to understand if this is the best annotation for this gene in *D. virilis,* a comparison between the annotations of its ortologous in the other twelve species was made. As we can see in table 5, the annotation through these twelve *Drosophila* species is very unstable. For that reason we searched for the structure of the protein codified by this gene in *D. melanogaster* and it does not show the Mth typical features. Since this gene is one of the eldest genes of this family and despite all these miss annotation problems the annotation of this gene in *D. virilis* genome was used as a model to annotate the ortholog present in *D. americana* genome.

Table 4: Comparative study concerning the annotation of *mth*-like genes present in *D. virilis* genome and their orthologous in *D. melanogaster*. bp, base pair.

mth-like	*D. mel* (*CG*) *vs.* *D. vir.* (*GJ*)	CDS size (bp)	Number of Exons	Size of exons (bp, 5'➔3')
1	***CG4521***	2031	1	2031
	GJ14760	2040	1	2040
	–	–	–	–
	GJ16328	1308	5	440, 138, 206, 263, 256
5	***CG6965***	1491	6	252, 124, 170, 181, 500, 258
	GJ23397	1470	6	231, 127, 170, 178, 500, 258
11	***CG31147***	1533	5	494, 384, 260, 168, 222
		1482	6	494, 147, 185, 260, 168, 22
	GJ23561	1404	6	434, 147, 188, 263, 168, 198
14	***CG32476***	1602	7	115, 419, 189, 423, 76, 118, 255
	GJ16090	1380	5	292, 687, 133, 85, 178
8	***CG32475***	1479	6	503, 156, 191, 218, 357, 48
	GJ12496	1509	6	530, 156, 191, 218, 363, 45
9	***CG17084)***	1542	7	639, 191, 233, 156, 203, 79, 34
	GJ12495	1554	7	651, 191, 233, 156, 203, 79, 34
10	***CG17061***	1734	4	753, 612, 194, 171
		1758	4	753, 612, 224, 165
		1755	4	753, 612, 224, 162
		1674	4	753, 612, 194, 111
		1728	4	753, 612, 194, 165
	GJ12492	1626	5	738, 88, 433, 194, 168
	–	–	–	–
	GJ12490	2556	12	299, 198, 285, 204, 479, 159, 183, 266, 82, 88, 194, 108

Table 5: Comparison of the annotation of *mth*-like 14 in the twelve species of *Drosophila*, whose genome is publicaly available. bp, base pair.

Drosophila species	CDS size (bp)	Number of Exons	Size of exons (bp, 5'→3')
D. simulans	1605	7	114, 422, 189, 423, 76, 118, 255
D. sechellia	1605	7	115, 422, 189, 423, 76, 118, 255
D. melanogaster	1602	7	115, 419, 189, 423, 76, 118, 255
D. yakuba	1677	7	115, 494, 189, 423, 76, 118, 255
D. erecta	1590	7	115, 413, 189, 423, 76, 112, 255
D. ananassae	1590	6	100, 401, 189, 500, 133, 261
D. pseudoobscura	1200	5	124, 186, 500, 124, 261
D. persimilis	1200	5	124, 189, 500, 117, 261
D. willistoni	756	5	124, 183, 50, 127, 267
D. mojavensis	1509	6	19, 401, 687, 133, 85, 178
D. virilis	1380	5	292, 687, 133, 85, 178
D. grimshawi	1425	5	337, 687, 133, 89, 178

3.2.2. Annotation of *mth*-like genes in *D. americana* and searching for conserved domains

In order to design specific molecular markers for each *mth*-like gene present in *D. americana* genome, firstly it was necessary to annotate all these genes in the genome of both strains. After annotating, all the *mth*-like genes present in *D. americana*, with the help of Conserved Domains tool at NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), it was possible to identify which conserved domains are codified in these genes. From this study, we were able to predict that *D. americana* has three *mth*-like genes without Mth ectodomain, and six which have the Mth ectodomain. However, as we can see on table 6, *mth*-like 8 and *mth*-like 9 have an Mth ectodomain but do not show a 7tm domain. Instead of a 7tm domain, *mth*-like 8 has a wcaJ_sugtrans domain (undecaprenyl-phosphate glucose phosphotransferase). Besides that, *mth*-like 1 seems to have only a 7tm domain. The latter belongs to the group of the oldest

mth-like genes, which also comprises two more members that do not show any conserved domains – *mth*-like 5 and *mth*-like 14. Finally, we can say that *D. americana* encodes four *mth*-like genes that have the Mth typical features, in other words Mth ectodomain and a 7tm domain – orthologous of *GJ16328* and *GJ12490, mth*-like 11 and *mth*-like 10.

Table 6: Conserved domains encoded by *mth*-like genes of *D. americana*.

mth-like genes	Conserved Domains
***mth*-like 1**	7tm2
***GJ16328* ortholog**	Mth and 7tm2
***mth*-like 5**	–
***mth*-like 11**	Mth and 7tm2
***mth*-like 14**	–
***mth*-like 8**	Mth and wcaJ_sugtrans
***mth*-like 9**	Mth
***mth*-like 10**	Mth and 7tm2
***GJ12490* ortholog**	Mth, 7tm2 and 7tm6

3.3. Lifespan variation in *D. americana*

3.3.1. Molecular markers design

Then it was found out all amino acid and nucleotide polymorphisms as well as differences in introns size existing between both *D. americana* strains (H5 and W11) in the same gene. After this identification, a molecular marker that distinguishes the polymorphism selected for each *mth*-like gene was designed. It was given preference for molecular markers that distinguish amino acid polymorphisms, since a change of an amino acid can interfere with the normal function of a protein. It was possible to design molecular markers for all *mth*-like genes present in *D. americana* (table 7). However it was not ruled out the possibility that the influence of these genes on lifespan is due to differences in the gene expression and not just due to a polymorphism.

Table 7: Localization of *mth*-like genes in *D. americana* genome and respective type of marker.

mth-like genes	Chromosome	Type of Marker
***mth*-like 1**	*X*	amino acid
***GJ16328* ortholog**	*2*	amino acid
***mth*-like 5**	*2*	nucleotide
***mth*-like 11**	*2*	intron size difference
***mth*-like 14**	*3*	amino acid
***mth*-like 8**	*3*	amino acid
***mth*-like 9**	*3*	nucleotide
***mth*-like 10**	*3*	intron size difference
***GJ12490* ortholog**	*3*	nucleotide

3.3.2. Lifespan of individuals from an F2 association study

All the markers listed above were then tested in all individuals from the first F2 association study previously described. This F2 association study comprises five crosses among five different strains from the species in study – *D. americana* – with the aim to test if there is association between the different *mth*-like genes and lifespan variation.

As previously stated, *D. americana* is a native species of United States and has a large effective population size. This species can be found in the central and eastern regions of United States from the south to the north, and this influenced the choice of strains used. The strains W11 and H5 were chosen since their genomes were recently sequenced. The other three species were selected in order to reproduce in this experiment what happens in the natural environment. So a strain typical from the north of distribution (O57) as well as two strains characteristic from the south of distribution (W29 and W46) were chosen, in order to include a high number of individuals and most of the relevant variation that exists at the molecular level.

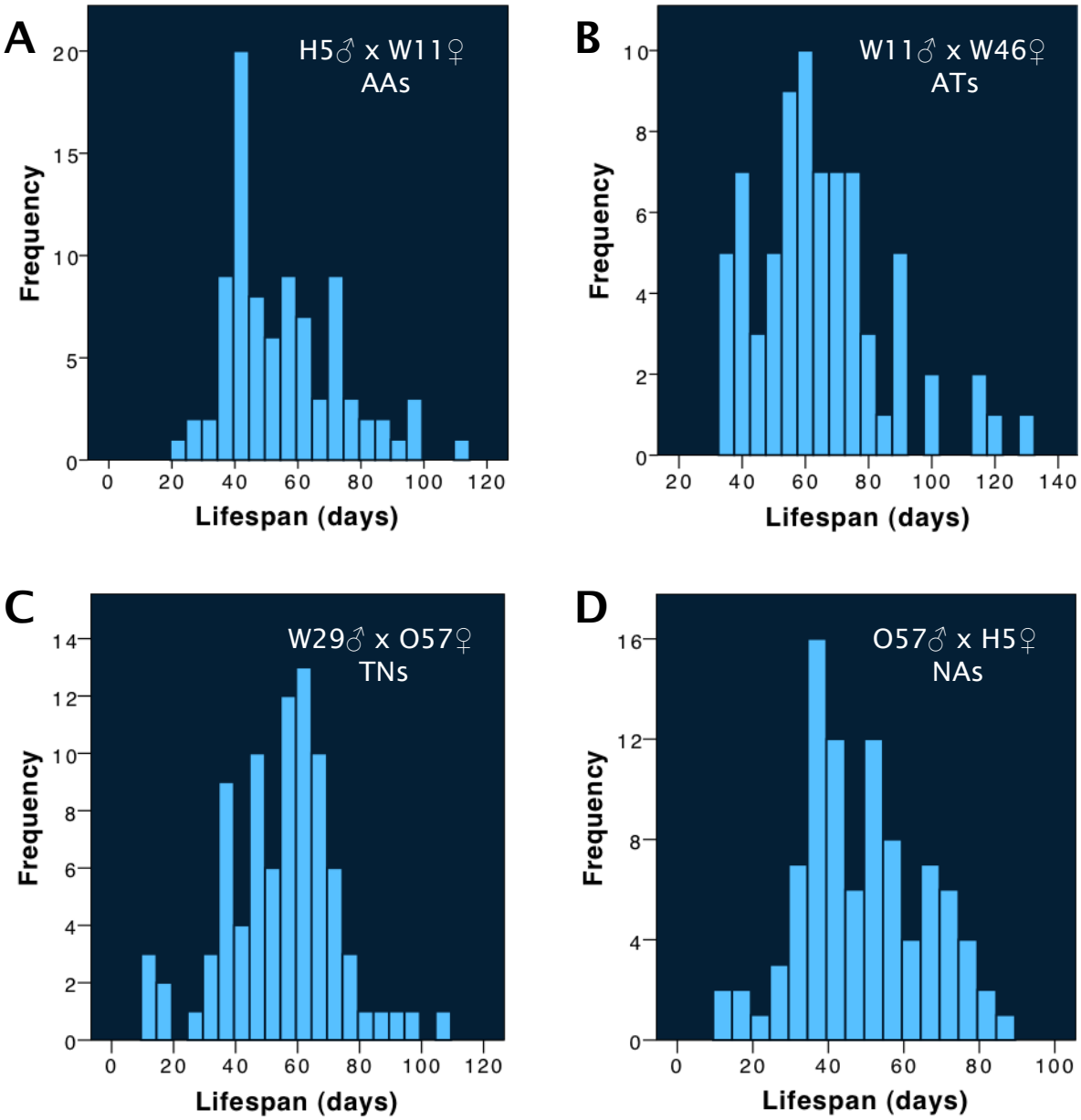

Figure 9: Lifespan variation observed in F2 individuals of each cross of the F2 association study. (A) cross between H5♂ x W11♀ (AAs), Mean=55,06 days, Standard Deviation=18,27, N=88; (B) cross between W11♂ x W46♀ (ATs), Mean=64,88 days, Standard Deviation =20,61, N=75; (C) cross between W49♂ x O57♀ (TNs), Mean=54,59 days, Standard Deviation =17,86, N=87; (D) cross between O57♂ x H5♀ (NAs), Mean=49,01 days, Standard Deviation =16,49, N=93; (E) cross between W46♂ x W29♀ (TTs), Mean=46,47 days, Standard Deviation =17.24. N=89.

As we can see on figure 9, the number of individuals varies according to the crosses, as well as the average lifespan of individuals from each cross. And since there is variation in phenotype, there should also be some variation in the genotype that justifies the difference in longevity.

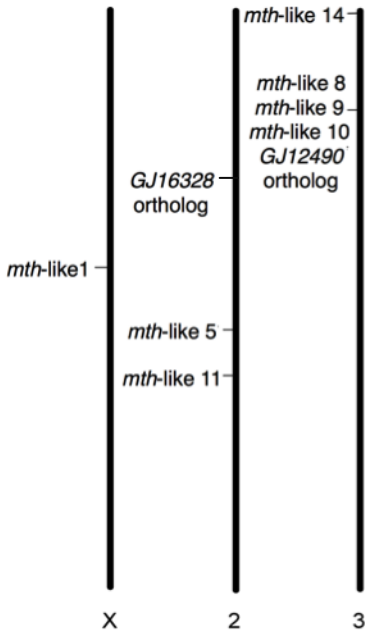

Figure 10: Chromosomal localization of *mth*-like genes identified in *D. americana.*

In the *D. americana* F2 association study there is a single generation of recombination, and thus closely linked markers will give similar information. The localization of all *mth*-like genes present in *D. americana* genome is shown in figure 10 and from this it is possible to see that there is a cluster of four *mth*-like genes in chromosome 3 (Muller's element E). Moreover, there is a *mth*-like gene localized in chromosome *X* (Muller's element A) – *mth*-like 1 -, and once only males were tested, there is only one allele present in each individual and there are not heterozygous individuals. The markers were firstly typed on the ten F0 individuals and in the crosses where the marker was segregating then the F2 individuals were genotyped.

For the *mth*-like 1 gene, that is localized in chromosome *X,* the marker was segregating in two crosses: AA and AT. After digesting the amplification product with *Msp*I, two different genotypes were found in these crosses, namely homozygous 0 (no digestion) and homozygous 1 (fully digested). In AA cross, the F0 male was homozygous 1 and the female was homozygous 0 and in AT cross, as expected F0 male was homozygous 0 and female was homozygous 1. In the F2 association experiment, AA individuals homozygous 1 (49.3±2.75 days; N=44) and homozygous 0 (61.4±2.51 days; N=43) show a significant association with lifespan (Non-parametric Kruskal-Wallis Test; P=0.000). In the individuals originated from AT cross, homozygous 1 (64.0±4.60; N=22) and homozygous 0 (66.1±3.11 days; N=43) were found, but no association was observed with lifespan (Non-parametric Kruskal-Wallis Test; P=0.708). Therefore, genotype 1

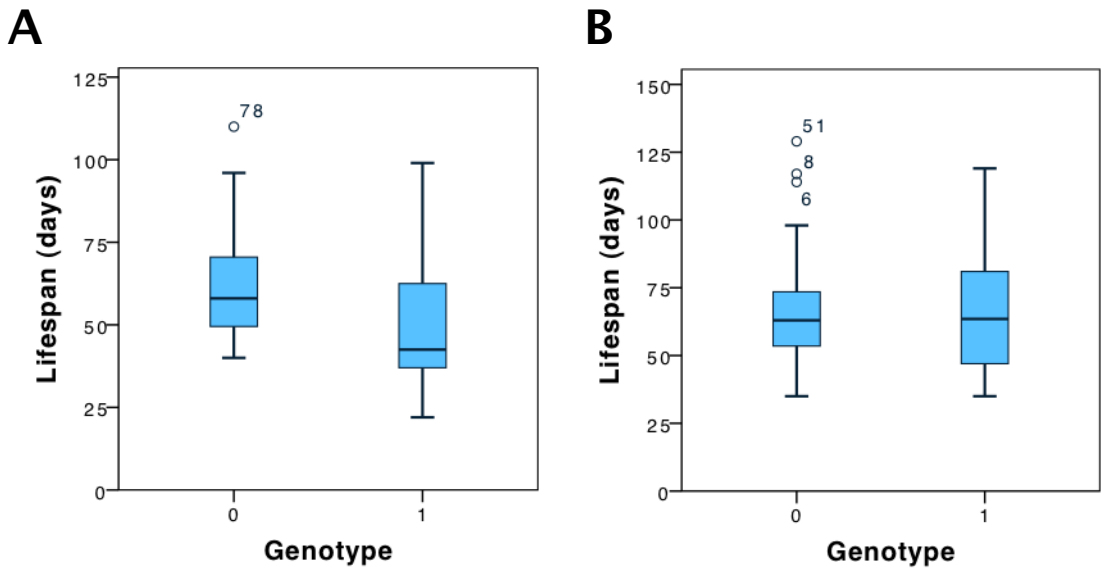

Figure 11: Box plot representing the association results obtained for gene *mth*-like 1 in F2 individuals from cross AA (A) and cross AT (B). The top of the box represents the 75th percentile, the bottom of the box represents the 25th percentile, and the line in the middle represents the 50th percentile that corresponds to the median. The whiskers represent the highest and lowest values that are not outliers or extreme values. Circles beyond the whiskers represent outliers

from H5 seems to be associated with short lifespan and genotype 0 from W11 seems to be associated with normal lifespan (figure 11). Moreover, allele 1 from W46 could have differences in comparison with allele 1 from H5, once in AT cross, individuals homozygous for allele 1 lived about more 15 days than individuals homozygous for allele 1 in AA cross.

The amplification product of orthologous gene of *GJ16328* after the digestion with *Hyp*CH4IV in F0 individuals showed two different genotypes: homozygous 1 (fully digested) and heterozygous for allele 1 and 0 (half-digested). However, only two out of the five crosses were segregating the marker: AA and AT. The AA male was homozygous for allele 1 and the AA female was heterozygous for allele 1 and 0 (1/0). On the other hand, AT male was heterozygous 1/0 and the female was homozygous for allele 1. The other six F0 individuals were homozygous for allele 1. The F2 descendent individuals from AA and AT crosses were then genotyped for this marker. Three genotypes were found in F2 individuals from AA cross: homozygous for allele 0 (60.8±4.03 days; N=16), homozygous for allele 1 (44.6±3.42 days; N=16) and heterozygous 1/0 (57.3±2.64 days; N=53) and an association between genotype and lifespan was found in this cross (Non-

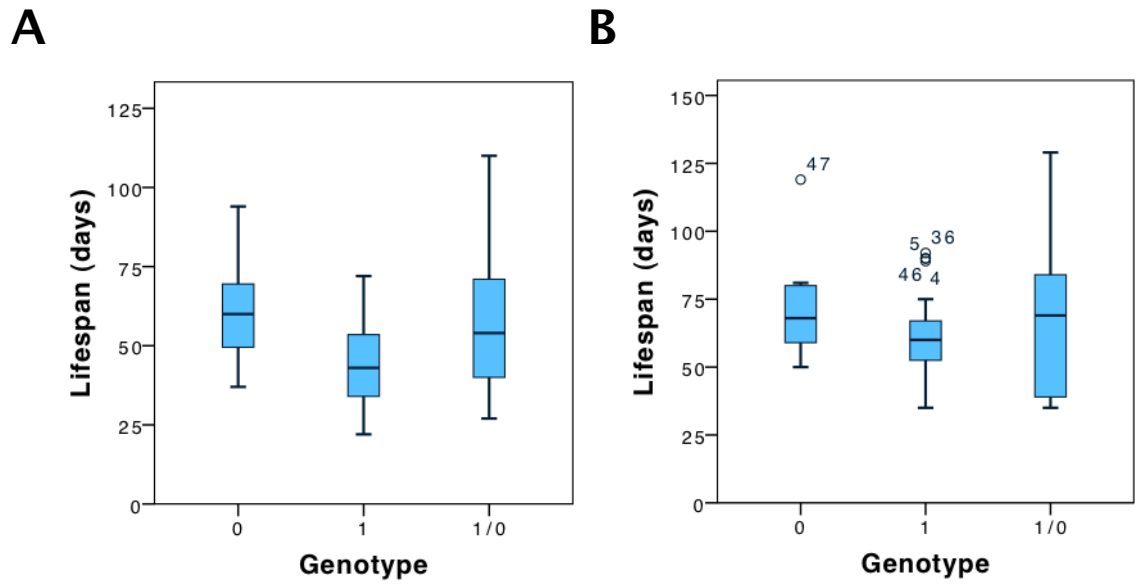

Figure 12: Box plot representing the association results obtained for *GJ16328* ortholog in F2 individuals from cross AA (A) and cross AT (B). The top of the box represents the 75th percentile, the bottom of the box represents the 25th percentile, and the line in the middle represents the 50th percentile that corresponds to the median. The whiskers represent the highest and lowest values that are not outliers or extreme values. Circles beyond the whiskers represent outliers.

parametric Kruskal-Wallis Test; P=0.028). The individuals originated from AT cross also showed three different genotypes: homozygous for allele 0 (70.9±4.66 days; N=14), homozygous for allele 1 (60.7±2.30 days; N=39) and heterozygous 1/0 (68.6±6.72 days; N=17), but no association was found between genotype and lifespan in this cross (Non-parametric Kruskal-Wallis Test; P=0.148). With these crosses, it is not possible to define which allele is dominant, once there is a common allele in the individuals from F0. However, it is clear that the genotype homozygous 1 is associated with short lifespan in both crosses (figure 12). Individuals homozygous 1 from AA lived about 16 days less than homozygous 0 and 13 days less than heterozygous 1/0 from the same cross. In the F2 individuals originated from AT cross we can see the same behaviour: homozygous 1 lived approximatly 11 days less than homozygous 0 and 8 days less than heterozygous 1/0. From these results, besides concluding that allele 1 is associated with short lifespan, we are able to say that allele 0 balances the effect of allele 1 once the heterezygous 1/0 show an average lifespan similar to the homozygous 0 in both crosses.

For *mth*-like 5 the marker is segregating in three crosses – AA, TN and NA. After digesting the amplification product with *Bcl*I three genotypes were observed: homozygous 0 (no digestion), homozygous 1 (fully digested) and heterozygous for allele 0 and 1 (1/0). However, the F0 individuals only showed two of the three genotypes predicted: in AA cross, the male was heterozygous 1/0 and the female was homozygous 0; in TN cross, the male was homozygous 0 and the female was heterozygous 1/0; and in NA cross, the male was homozygous 0 and the female was heterozygous 1/0. F2 individuals from AA cross were homozygous 0 (57.8±2.76 days; N=47), only one homozygous 1 (43 days), and heterozygous 1/0 (52.43±2.95 days; N=37) but no association with lifespan was found (Non-parametric Kruskal-Wallis Test; P=0.489). In F2 individuals derived from TN cross

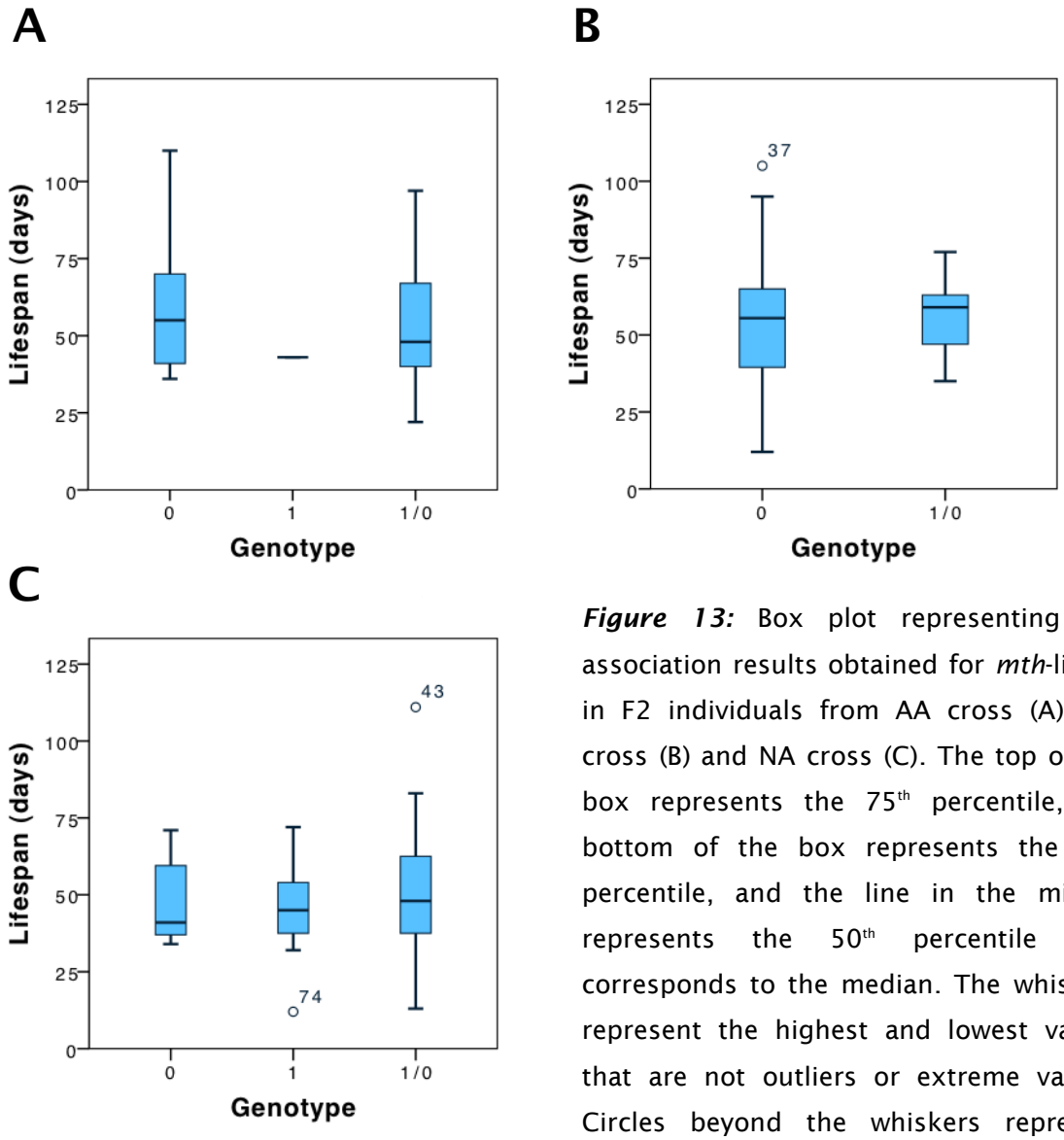

Figure 13: Box plot representing the association results obtained for *mth*-like 5 in F2 individuals from AA cross (A), TN cross (B) and NA cross (C). The top of the box represents the 75th percentile, the bottom of the box represents the 25th percentile, and the line in the middle represents the 50th percentile that corresponds to the median. The whiskers represent the highest and lowest values that are not outliers or extreme values. Circles beyond the whiskers represent outliers.

no homozygous 1 was found. Although, it was possible to find homozygous 0 (53.4±2.58 days; N=60) and heterozygous 1/0 (56.4±2.35 days; N=25) individuals and no association between genotype and lifespan was found (Non-parametric Kruskal-Wallis Test; P=0.458). In the individuals originated from the last cross – NA – an association between this genotype and lifespan was not also found (Non-parametric Kruskal-Wallis Test; P=0.742). In the descendents of this cross it was possible to find homozygous 0 (48.3±4.09 days; N=11), homozygous 1 (46.1±4.01 days; N=15) and heterozygous 1/0 (50.4±2.30 days; N=64). It seems that no matter what allele an individual has, the mean longevity for each genotype in each cross is very similar (figure 13). However, it is noted that individuals homozygous 1 are very rare.

The marker for *mth*-like 11 is the only one that is segregating in all five crosses. There is, however, only one cross where an association was found between genotype and lifespan - AA cross (Non-parametric Kruskal-Wallis Test; AA cross P=0.003; AT cross P=0.452; TN cross P=0.173; NA cross P=0.444; and TT cross P=0.353). This marker showed fourteen types of genotypes in F2 individuals: homozygous 1, homozygous 2, homozygous 3, homozygous 4, homozygous 5 and heterozygous 1/2, heterozygous 1/3, heterozygous 1/4, heterozygous 2/3, heterozygous 2/4, heterozygous 2/5, heterozygous 3/4, heterozygous 3/5 and heterozygous 4/5. In the AA cross, the F0 male was heterozygous 2/4 and the female heterozygous 3/4, so allele size 4 was shared by both strains. Thus, in the F2 individuals from this cross we found five out of the six possible genotypes, once no homozygous 2 was detected: homozygous 3 (61.6±4.99 days; N=13), homozygous 4 (56.6±4.79 days; N=9), heterozygous 2/3 (52.4±2.96 days; N=19), heterezygous 2/4 (37.9±2.60 days; N=12) and heterozygous 3/4 (61.0±3.92 days; N=30). In the founder members of AT cross, the male was heterozygous 3/4 and the female heterozygous 1/2, so in the F2 individuals we found ten different genotypes: homozygous 1 (60 days; N=1), homozygous 2 (52.0±10 days; N=2), homozygous 3 (82.3±18.48 days; N=3), homozygous 4 (68.7±5.74 days; N=14), heterozygous 1/2 (56.9±2.40 days; N=10), heterozygous 1/3 (61.6±11.00 days; N=8), heterozygous 1/4 (63.7±5.91 days; N=9), heterozygous 2/3 (72.2±5.86 days; N=12), heterozygous 2/4 (60.4±7.04 days; N=7) and heterozygous 3/4 (75.5±12.9; N=4). In the F0 individuals of TN cross, the male was homozygous 2 and the female was heterozygous 3/5, thus in the F2

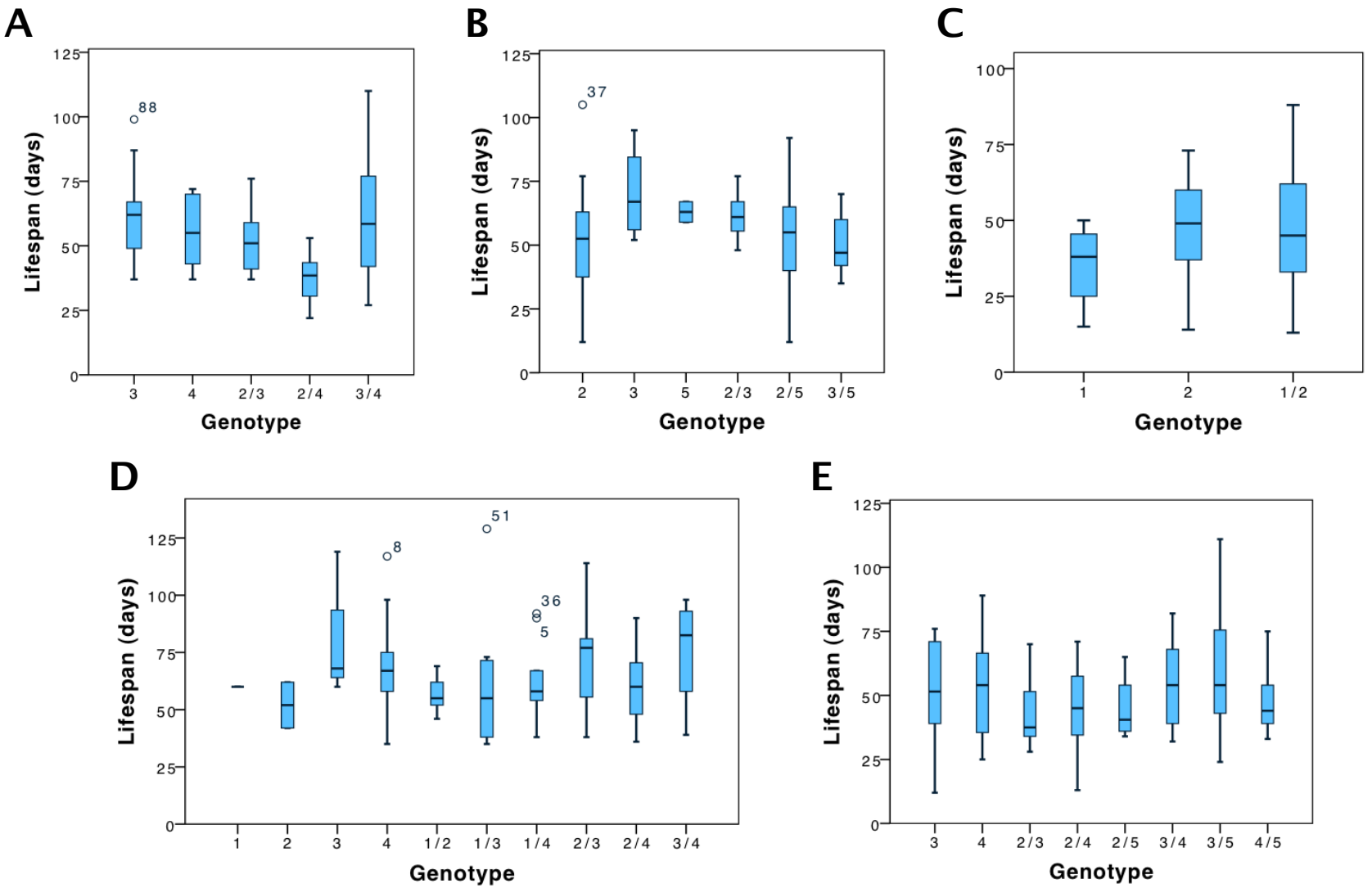

Figure 14: Box plot representing the association results obtained for *mth*-like 11 in F2 individuals from AA cross (A), TN cross (B), TT cross (C), AT cross (D) and NA cross (E). The top of the box represents the 75th percentile, the bottom of the box represents the 25th percentile, and the line in the middle represents the 50th percentile that corresponds to the median. The whiskers represent the highest and lowest values that are not outliers or extreme values. Circles beyond the whiskers represent outliers.

individuals six different genotypes were found: homozygous 2 (50.9±4.96 days; N=20), homozygous 3 (70.2±9.41 days; N=4), homozygous 5 (63.0±4.00 days; N=2), heterozyous 2/3 (61.8±2.48 days; N=12), heterozygous 2/5 (53.1±3.32 days; N=33) and heterozygous 3/5 (51.0±3.15 days; N=13). In NA cross the male was heterozygous 3/5 and the female heterozygous 2/4, so in the F2 individuals we found eight of the ten possible genotypes, once neither homozygous 2 nor homozygous 5 were found. So, in the F2 individuals from TN cross there were homozygous 3 (51.4±7.61 days; N=8), homozygous 4 (53.2±7.56 days; N=8), heterozygous 2/3 (43.0±4.91 days; N=8), heterozygous 2/4 (44.5±4.50 days; N=16), heterozygous 2/5 (45.0±4.00; N=8), heterozygous 3/4 (54.8±3.47 days; N=21), heterozygous 3/5 (60.0±9.62; N=8) and heterozygous 4/5 (47.5±3.48 days; N=13). Finally in the TT cross, the male was heterozygous 1/2 and the female was homozygous 2. So, in the F2 individuals there were three genotypes possible: homozygous 1 (35.2±7.42 days; N=4), homozygous 2 (46.5±2.10 days; N=57) and heterozygous 1/2 (47.4±4.20 days; N=21).

Although an association with lifespan was found in AA cross only, there are some other conclusions that we can draw. For instance, in figure 14 it is possible to see that genotypes involving allele size 2 tend to live less than the other three (homozygous 3, homozygous 4 and heterozygous 3/4). Allele size 2 seems to be associated with short lifespan and to be dominant over allele 3 and 4, however, the latter two alleles do not seem to be equivalent once heterozygous 2/3 live on average 15 days more than heterozygous 2/4. Moreover, when homozygous 2 are absent in either AA and NA crosses, allele 2 has the same origin in both crosses – H5 strain. So, it seems likely that once it has the same origin in both crosses, this allele could be lethal, so no homozygous 2 are found in F2 indivudals of these crosses.

For the *mth*-like 14 the F0 individuals showed two different genotypes homozygous 0 (not digested) and homozygous 1 (fully digested). This marker was segregating in four crosses: AA, AT, NA and TT. The male from F0 of AA cross was homozygous 1 and the female was homozygous 0. Therefore, in the F2 individuals it was possible to find three different genotypes: homozygous 0 (57.4±3.21 days; N=29), homozygous 1 (52.0±3.18 days; N=13) and heterozygous 1/0 (53.6±3.00 days; N=44). The AT cross founders were also homozygous for one of the two alleles – male W11 was homozygous 0 and

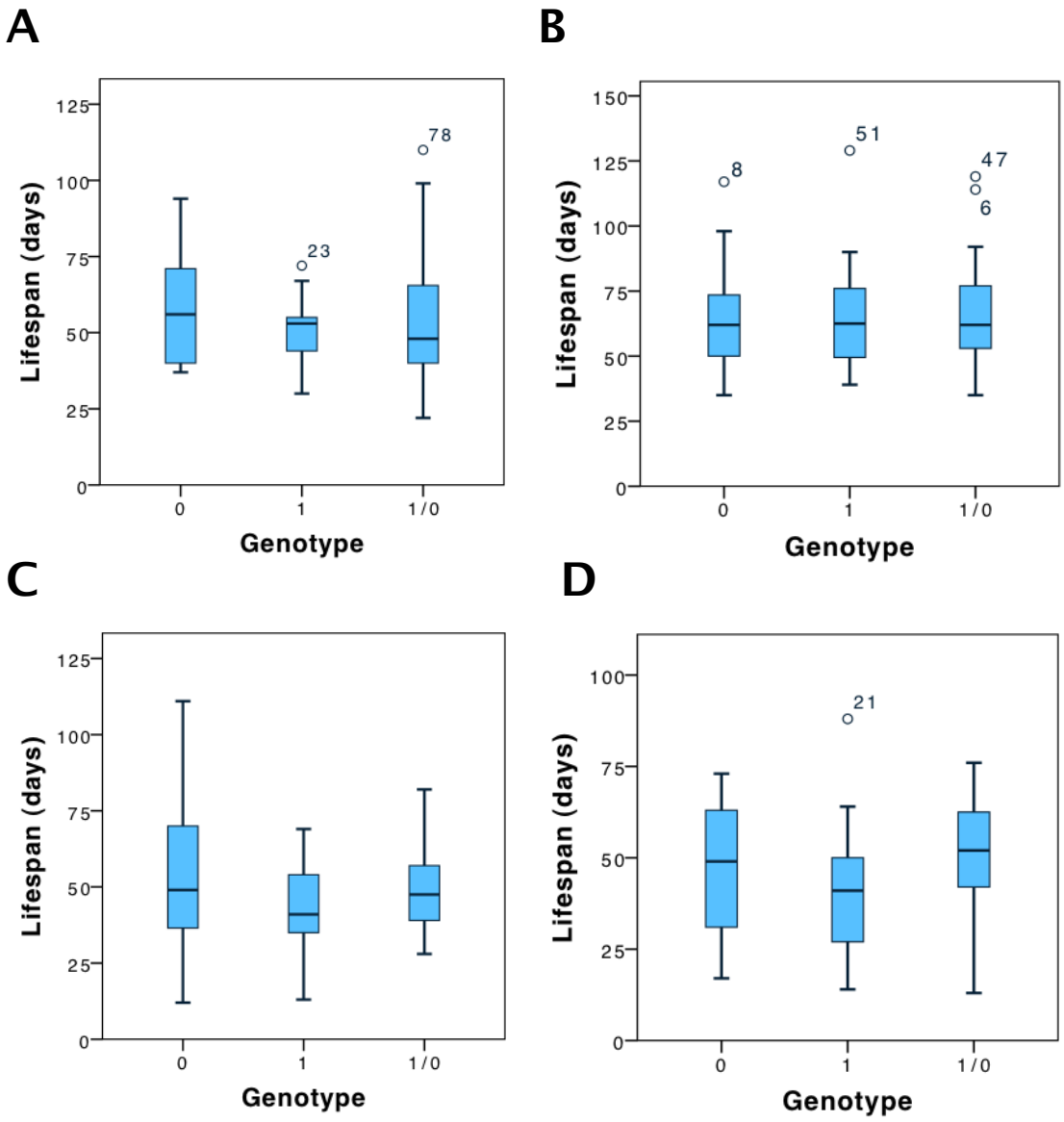

Figure 15: Box plot representing the association results obtained for *mth*-like 14 in F2 individuals from AA cross (A), AT cross (B), NA cross (C) and TT cross (D). The top of the box represents the 75th percentile, the bottom of the box represents the 25th percentile, and the line in the middle represents the 50th percentile that corresponds to the median. The whiskers represent the highest and lowest values that are not outliers or extreme values. Circles beyond the whiskers represent outliers.

female W46 was homozygous 1 – and the F2 individuals also showed three genotypes: homozygous 0 (63.8±4.31 days; N=24), homozygous 1 (65.8±5.30 days; N=16) and heterozygous 1/0 (66.3±3.47 days; N=33). In cross NA, the F0 male was homozygous 0 and the female homozygous 1, so in the F2 individuals there were homozygous 0 (52.4±3.58 days; N=36), homozygous 1 (44.0±4.08 days; N=17) and heterozygous 1/0 (49.6±2.13 days; N=38). In the last cross, the

F0 TT male was homozygous 1 and the female was homozygous 0, and the F2 presented the three types of possible genotypes: homozygous 0 (46.8±3.04 days; N=33), homozygous 1 (40.5±3.94 days; N=21) and heterozygous 1/0 (49.7±2.71 days; N=35). Comparing the mean of lifespan for the three genotypes in each cross no major differences are detected and this is supported by the fact that no association between lifespan and genotype was detected in none of the four F2 crosses (Non-parametric Kruskal-Wallis Test; AA cross P=0.590; AT cross P=0.872; NA cross P=0.466; TT cross P=0.093) (figure 15).

In relation to *mth*-like 8, its marker was also segregating in four out of the five F2 crosses: AA, AT, TN and NA. After the digestion of amplification product with *Sau*3AI there were three genotypes: homozygous 0 (not digested), homozygous 1 (fully digested) and heterozygous 1/0 (half-digested). In the founder members of these four crosses one member was homozygous 0 and the other was homozygous 1: AA male homozygous 1 and AA female homozygous 0; AT male homozygous 0 and AT female homozygous 1; TN male homozygous 1 and TN female homozygous 0; and NA male homozygous 0 and NA female homozygous 1. In the F2 individuals descendents of AA cross there were homozygous 0 (60.7±3.69 days; N=25), homozygous 1 (44.9±3.38 days; N=18) and heterozygous 1/0 (56.4±2.77 days; N=44) individuals. In this cross a strong association with lifespan was found (Non-parametric Kruskal-Wallis Test; P=0.013) and this is supported by the fact that homozygous 0 and heterozygous 1/0 lived about, respectively, 16 days and 12 days longer than homozygous 1. This also suggests that genotype 0 can be dominant over genotype 1 and when in heterozygosity the allele 0 masks the effect of allele 1, which is probably associated with short lifespan (figure 16 A). In AT cross no association between genotype and lifespan was found (Non-parametric Kruskal-Wallis Test; P=0.088), once the mean lifespan of the F2 individuals of each genotype were similar: homozygous 0 (77.0±6.96 days N=18), homozygous 1 (64±4.42 days; N=23) and heterozygous 1/0 (62.3±2.74 days; N=37). The TN F2 individuals also did not show association between genotype and lifespan (Non-parametric Kruskal-Wallis Test; P=0.998), once all genotypes have a similar average lifespan: homozygous 0 (55.6±2.83 days; N=32), homozygous 1 (54.8±4.22 days; N=21) and heterozygous 1/0 (52.9±3.30 days; N=33). However, in the NA F2 individuals, a strong association between genotype and lifespan was found (Non-parametric Kruskal-Wallis Test; P=0.001). The homozygous 0 (62.4±3.91 days; N=23) lived

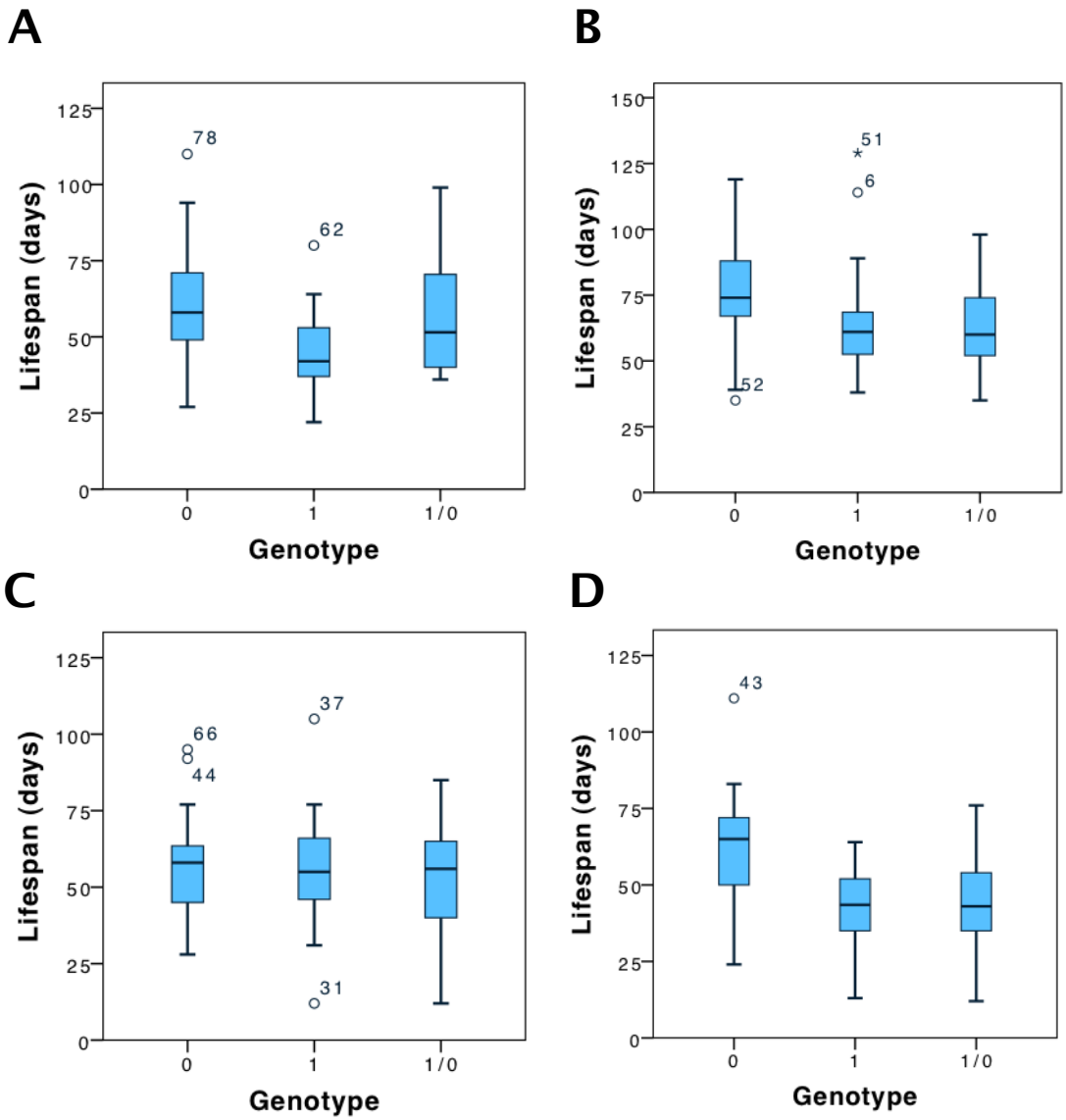

Figure 16: Box plot representing the association results obtained for *mth*-like 8 in F2 individuals from AA cross (A), AT cross (B), TN cross (C) and NA cross (D). The top of the box represents the 75th percentile, the bottom of the box represents the 25th percentile, and the line in the middle represents the 50th percentile that corresponds to the median. The whiskers represent the highest and lowest values that are not outliers or extreme values. Circles beyond the whiskers represent outliers and star represents an extreme outlier.

about more 20 days than homozygous 1 (42.0±3.59 days; N=18) and 16 days than heterozygous 1/0 (46.4±2.04 days; N=49). As stated above, the results of this latter cross support the idea that genotype 1 seems to be associated with short lifespan. However, the idea that when in heterozygosity allele 0 masks the effect of allele 1 is not observed in NA F2 individuals. It is also important to retain

the fact that when there is association between genotype and lifespan, the allele 1 that seems to be associated with short lifespan comes from H5 strain (figure 16).

For *mth*-like 9, a strong association between genotype and lifespan was found in the two crosses where the marker was segregating – AA and NA (Non-parametric Kruskal-Wallis Test; P=0.012 and P=0.000, respectively). In AA cross, the male was homozygous 0 and the female was homozygous 1. So, in F2 individuals derived from this cross there were homozygous 0 (41.9±2.24 days; N=14), homozygous 1(58.7±3.48 days; N=28) and heterozygous 1/0 (56.0±2.80 days; N=43), and it is possible to see that homozygous 0 lived less than the other two possible genotypes. The NA cross founders were also homozygous for one of the two alleles: the male was homozygous 1 and the female was homozygous 0. So, F2 individuals from NA cross also showed these three genotypes: homozygous 0 (42.5±3.76 days; N=18), homozygous 1 (62.4±3.91 days; N=23) and heterozygous 1/0 (45.3±1.94 days; N=48). However, from this latter cross it is only possible to conclude that genotype 1 seems to be associated with longer

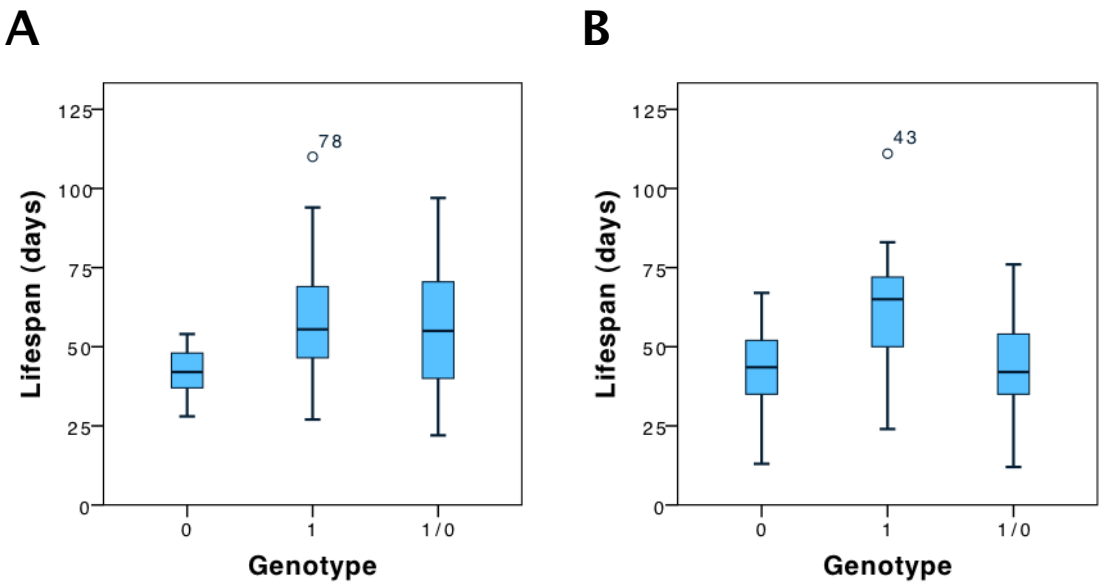

Figure 17: Box plot representing the association results obtained for *mth*-like 9 in F2 individuals from AA cross (A) and AT cross (B). The top of the box represents the 75th percentile, the bottom of the box represents the 25th percentile, and the line in the middle represents the 50th percentile that corresponds to the median. The whiskers represent the highest and lowest values that are not outliers or extreme values. Circles beyond the whiskers represent outliers.

lifespan. There are three conclusions that we can take from these results (figure 17): allele 0 seems to be associated with short lifespan once it comes from the same strain in both crosses (H5); allele 1 seems to be associated with longer lifespan, once homozygous 1 lived longer than homozygous 0; and finally it is not possible to say that allele 1 is dominant over allele 0, once in AA cross the average lifespan of heterozygous 1/0 is similar to the average lifespan of homozygous 1 of this cross (about 56 days), and in NA cross the average lifespan of heterozygous 1/0 is similar to the average lifespan of homozygous 0 (about 45 days).

For *mth*-like 10, when the F2 cross founding members were genotyped, three alleles of different sizes were found. However, this marker was only segregating in AA, TN and NA crosses. In AA cross, the F0 male was heterozygous for allele sizes 1 and 3 and the female was homozygous 2. Thus, in F2 individuals six different genotypes were found: homozygous 1 (39.0±2.00 days; N=2), homozygous 2 (60.4±4.36 days; N=22), homozygous 3 (43 days; N=1), heterozygous 1/2 (55.1±4.13 days; N=20), heterozygous 1/3 (40.9±3.68 days; N=8) and heterozygous 2/3 (58.3±3.21 days; N=30). No association was found in this cross between genotype and lifespan (Non-parametric Kruskal-Wallis Test; P=0.073). In the founders members of TN cross, the male was homozygous 2 and the female homozygous 1. So in the F2 individuals there were three genotypes: homozygous 1 (52.4±3.31 days; N=19), homozygous 2 (55.2±4.19 days; N=20) and heterozygous 1/2 (54.3±2.78; N=46) and also no association was found between genotype and lifespan in this cross (Non-parametric Kruskal-Wallis Test; P=0.622). However, an association between genotype and lifespan was found in the F2 individuals from NA cross (Non-parametric Kruskal-Wallis Test; P=0.001). The F0 male of this cross was homozygous 1 and the female was homozygous 3, therefore, in F2 individuals there were three genotypes: homozygous 1 (61.0±3.79 days; N=26), homozygous 3 (40.8±3.48 days; N=18) and heterozygous 1/3 (47.2±2.18 days; N=47). Despite of only one out of the three crosses where this marker was segregating showed association with lifespan, there is one conclusion we can draw: NA F2 individuals homozygous 3 and heterozygous 1/3 lived less than homozygous 1 in the same cross and, comparing with F2 AA individuals we are able to see that there is only one individual homozygous 3 (figure 18). Once in these two crosses the allele 3 is

originated from H5 strain, it seems that allele 3 from H5 is associated with short lifespan.

The amplification product of ortholog gene of *GJ12490* after the digestion with *Sau*3AI showed two different genotypes in F0 individuals: homozygous 0 (no digested) and homozygous 1 (fully digested). From the five crosses, only two

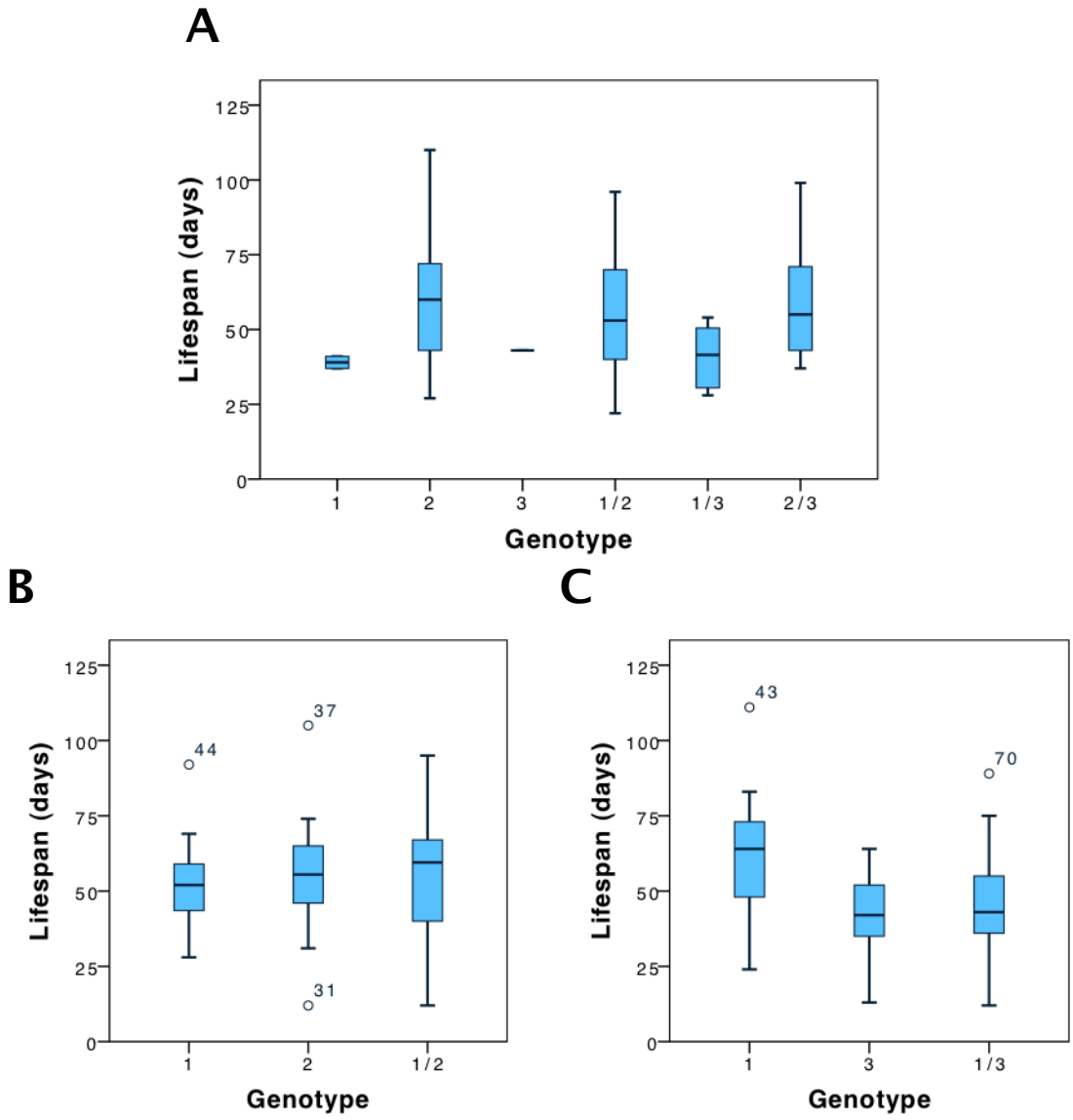

Figure 18: Box plot representing the association results obtained for *mth*-like 10 in F2 individuals from AA cross (A), TN cross (B) and NA cross (C). The top of the box represents the 75th percentile, the bottom of the box represents the 25th percentile, and the line in the middle represents the 50th percentile that corresponds to the median. The whiskers represent the highest and lowest values that are not outliers or extreme values. Circles beyond the whiskers represent outliers and star represents an extreme outlier.

segregate this marker: AA and AT crosses. F0 male of AA cross was homozygous 0 and the female was homozygous 1, so in the F2 individuals three different genotypes were found: homozygous 0 (42.0±2.54 days; N=12), homozygous 1 (57.7±3.24 days; N=31) and heterozygous 1/0 (57.1±2.90 days; N=44). In the founder members of AT cross, the male was homozygous 1 and the female homozygous 0, thus in F2 individuals there were homozygous 0 (62.2±4.38 days; N=17), homozygous 1 (76.7±5.90 days; N=16) and heterozygous 1/0 (62.5±3.00 days; N=39). An association between genotype and lifespan was found in the individuals from both crosses (Non-parametric Kruskal-Wallis Test; P=0.028 and P=0.049, respectively). The average lifespan is very similar between the homozygous 1 and heterozygous 1/0 in AA cross and they lived much longer than homozygous 0 (figure 19 A). Therefore, genotype 0 seems to be associated with short lifespan. However, when we look to the results of AT cross, the average lifespan is similar between homozygous 0 and heterozygous 1/0, who lived lesser than homozygous 1 (figure 19 B). From this cross we are able to conclude that genotype 1 seems to be related with longer lifespan. Nevertheless, when taken together no inference could be made in terms of allele dominance.

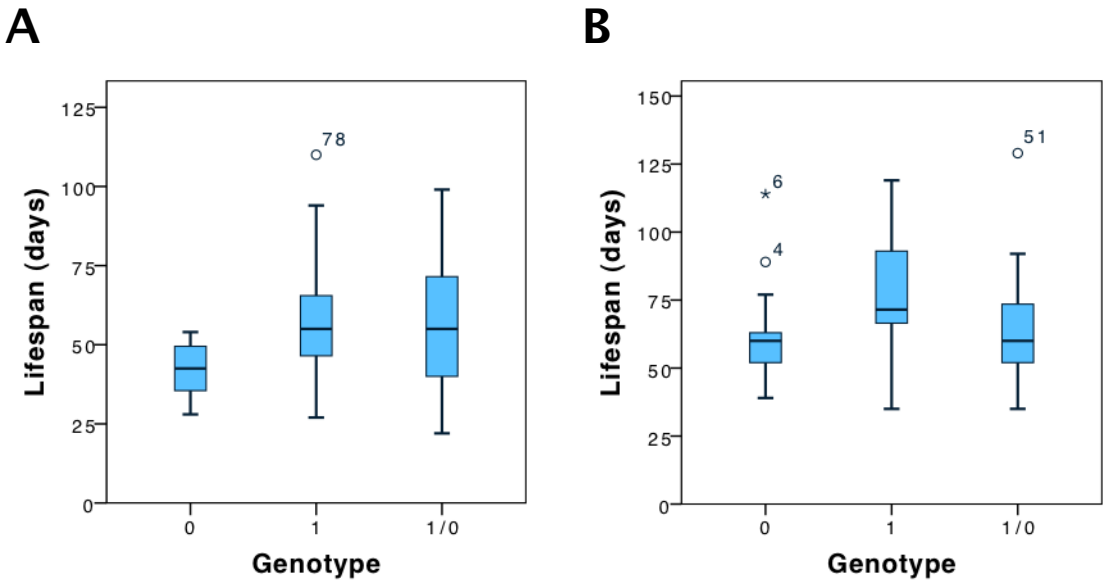

Figure 19: Box plot representing the association results obtained for *GJ12490* ortholog in F2 individuals from AA cross (A) and AT cross (B). The top of the box represents the 75th percentile, the bottom of the box represents the 25th percentile, and the line in the middle represents the 50th percentile that corresponds to the median. The whiskers represent the highest and lowest values that are not outliers or extreme values. Circles beyond the whiskers represent outliers and star represents an extreme outlier.

Grouping all the results, some more conclusions can be taken. As we can see on figure 20, in the genes that previously showed association with lifespan determination, the homozygous that carry the allele from H5 lived approximately 20 days less than the homozygous for the allele originated in the other founder strain. Moreover, the H5 allele seems to be recessive. This is especially notorious when we look for AA cross, where the allele from W11 seems to compensate the short lifespan determinated by allele from H5 strain (figure 20 B, F, G and I). On the other hand, in relation to genes that did not showed association with lifespan, these differences are not so notorious, once homozygous for the allele from H5, homozygous for the allele from the other founder strain and heterozygous lived approximately the same. Taking all this into account, we are able to say that the allele that came from H5 strain is in some way related to short lifespan. This means, that individuals that belong to H5 strain should live, in general, less than individuals from the other strains.

In order to try to understand why some genes show strong association than others it is essential to comprehend what differences are present in these proteins that might put in question their normal function. In figure 21, a scheme representative of the protein structure of each *mth*-like studied, as well as the amino acid differences between the two strains of *D. americana* – H5 and W11 – and the foreseen N-glycosylation sites are shown. The major difference is in the length of the proteins codified by these nine *mth*-like genes. *mth*-like 1 and the ortholog of *GJ12490* are much longer than the other genes. Moreover it is possible to see that the major part of amino acid polymorphisms occur inside Mth ectodomain, which is likely important for the protein to be functional, once the active site of this receptor is likely be here.

Another difference is the fact that ortholog of *GJ12490* has an additional 7tm domain (figure 21). Many efforts were done in order to understand if this 7tm domain is part of this gene or not, but no conclusion was reached. However, when a BLASTp is performed in NCBI (http://blast.ncbi.nlm.nih.gov/) using the sequence of the beginning of the gene till the end of the first 7tm domain as a query, there is a hit with odorant receptor of *D. pseudoobscura pseudoobscura*. So probably, this 7tm domain belongs to another gene that is next to *GJ12490* ortholog.

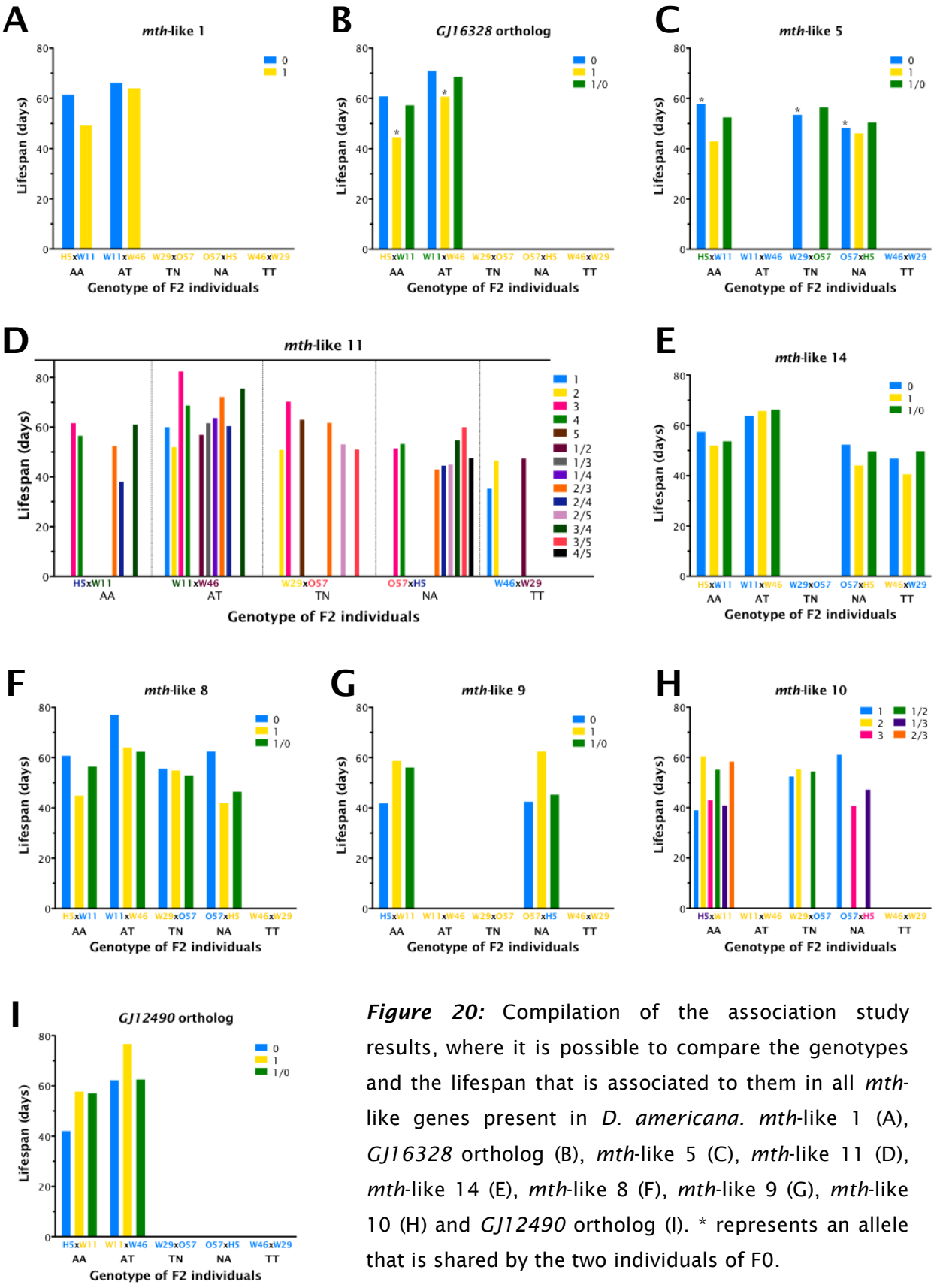

Figure 20: Compilation of the association study results, where it is possible to compare the genotypes and the lifespan that is associated to them in all *mth*-like genes present in *D. americana*. *mth*-like 1 (A), *GJ16328* ortholog (B), *mth*-like 5 (C), *mth*-like 11 (D), *mth*-like 14 (E), *mth*-like 8 (F), *mth*-like 9 (G), *mth*-like 10 (H) and *GJ12490* ortholog (I). * represents an allele that is shared by the two individuals of F0.

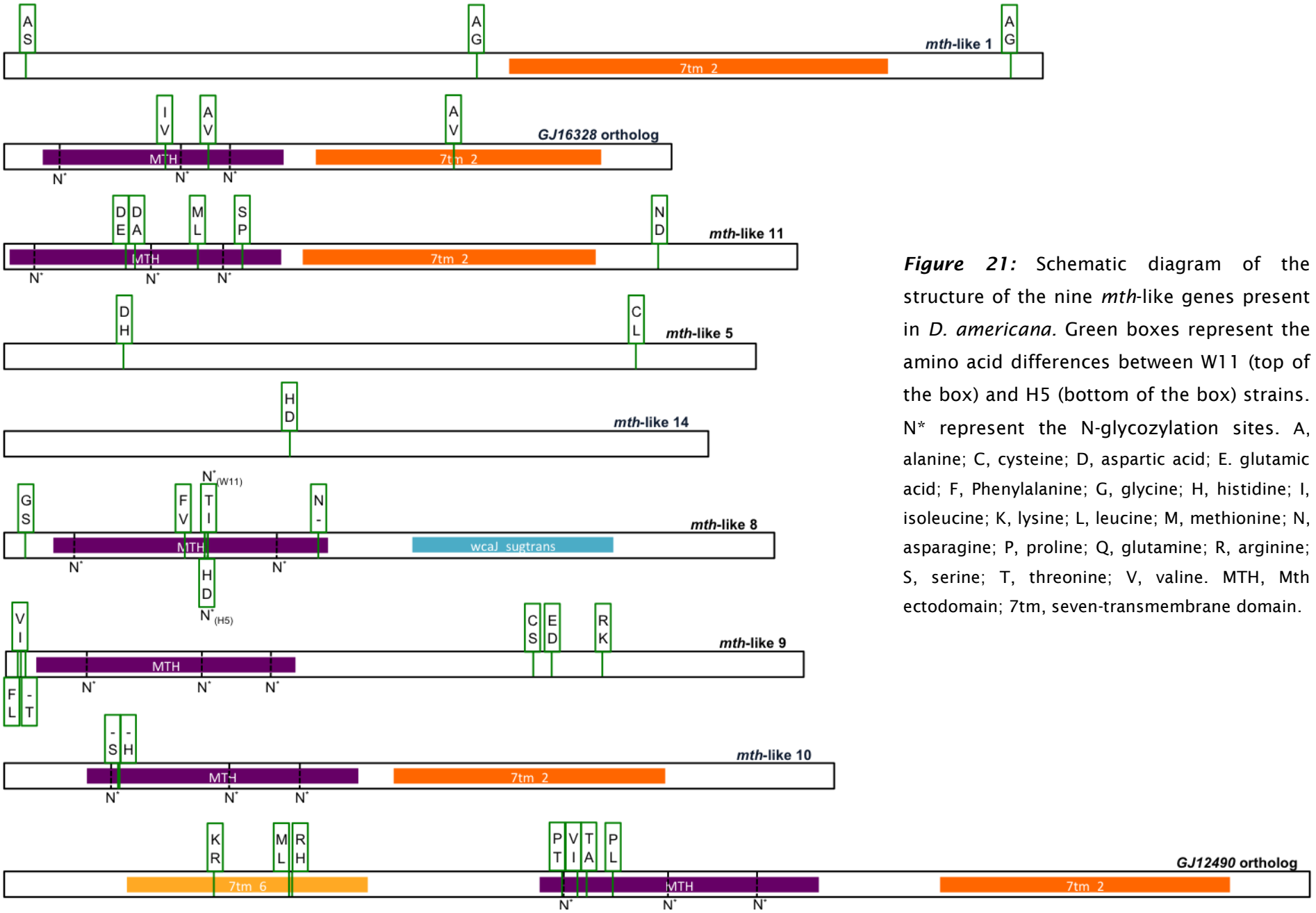

Figure 21: Schematic diagram of the structure of the nine *mth*-like genes present in *D. americana*. Green boxes represent the amino acid differences between W11 (top of the box) and H5 (bottom of the box) strains. N* represent the N-glycozylation sites. A, alanine; C, cysteine; D, aspartic acid; E. glutamic acid; F, Phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine. MTH, Mth ectodomain; 7tm, seven-transmembrane domain.

3.4. The second F2 association study

In order to test if the association between *mth*-like genes and lifespan determination remained untouched when the sample is bigger, another F2 association study was performed. The individuals of this F2 association experiment show a high diversity of phenotypes (figure 22), and for that reason some variation in genotypes might explain the variety in individuals' lifespan.

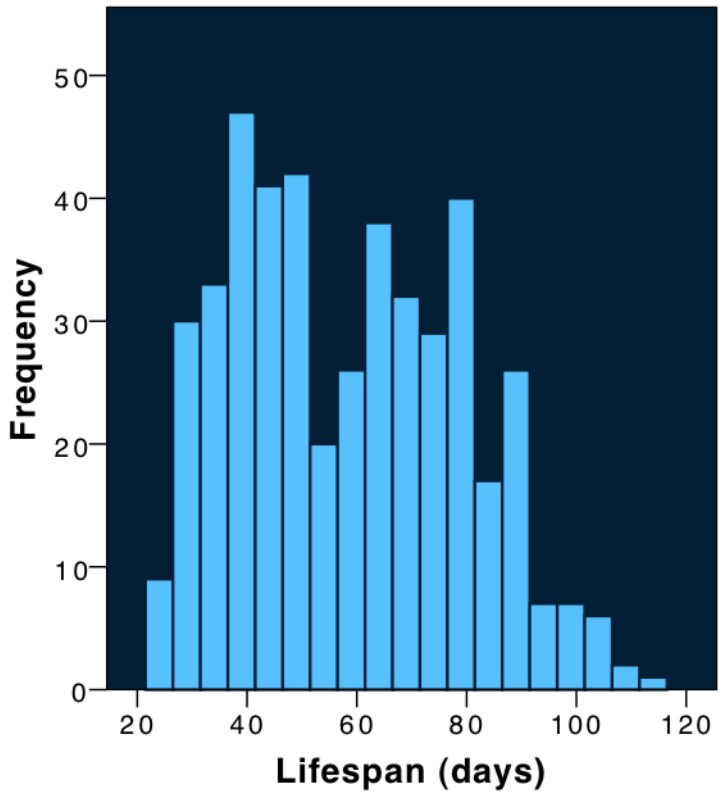

Figure 22: Lifespan variation observed in the 453 individuals of the second F2 association experiment (Mean=58.64 days, Standard Deviation=20.51, N=453).

In this F2 association study only three markers of *mth*-like genes of *D. americana* were tested. These genes were chosen due to its association or not with lifespan in the previous F2 association study. Firstly, the *mth*-like 5 was chosen in order to guarantee that when the sample is enlarged, *mth*-like genes without Mth ectodomain keep on showing no association with lifespan determination. Secondly, two *mth*-like genes, which show a strong association with lifespan, were chosen – ortholog of *GJ12490* and *mth*-like 11 – in order to confirm its association with lifespan. *mth*-like 5 and *mth*-like 11 are located in chromosome

2 (Muller's element E) and *GJ12490* ortholog is located in chromosome 3 (Muller's element D).

For *mth*-like 11, in this F2 association study, the size genotypes found were homozygous 3 (62.7 days±12.1 days; N=18) and homozygous 4 (64.2±19.8 days; N=173), and heterozygous for alleles size 2 and 3 (57.2±19.4 days; N=56), 2 and 4 (47.1±17.6 days; N=125), and 3 and 4 (67.2±21.6 days; N=65). Three alleles are present in this F2 association study because isofemale rather than isogenic strains were used. A significant association between this *mth*-like size genotype and lifespan was found (Non-parametric Kruskal-Wallis Test; $P<0.001$). When performing all possible pair wise allele size genotype comparisons, using Non-parametric Mann-Whitney tests, significant differences regarding lifespan are found for the following comparisons: 3 and 2/4 ($P<0.001$), 4 and 2/3 ($P<0.05$), 4 and 2/4 ($P<0.001$), 2/3 and 2/4 ($P<0.001$), 2/3 and 3/4 ($P<0.005$), and 2/4 and 3/4 ($P<0.001$). After applying the sequential Bonferroni correction for multiple testing, all significant comparisons with the exception of the one involving genotypes 4 and 2/3 are significant ($P<0.05$). These findings are not easily accounted for under a simple scenario, but we note that four out of the six most significant comparisons ($P<0.001$) involve genotype 2/4. Moreover, two out of the three comparisons involving genotype 2/3 are significant, although one of them is not significant after Bonferroni correction for multiple testing. These observations suggest that allele size 2 is associated with short lifespan and that is dominant over the alleles 3 and 4, though the latter two alleles are clearly not equivalent, because when comparing genotypes 2/3 and 2/4 it is possible to detect significant differences. On average, individuals that do not have allele size 2 live 28.8 % longer than individuals carrying allele size 2 (the average for individuals that do not have allele size 2 and for those that have allele size 2 is 64.9 and 50.4 days, respectively).

In *D. melanogaster*, more than 40 genes have been implicated in lifespan determination (Paaby and Schmidt, 2009). Therefore, the possibility that another gene in the vicinity of *mth*-like 11 is the one showing variation that influences lifespan must be taken into account. One of these candidate genes is *four wheel drive* (*fwd*) and this gene is in the middle of the genomic region defined by *mth*-like 11 and *mth*-like 5. For that reason, and under the hypothesis that variation at *fwd* rather than variation at *mth*-like 11 is the responsible for lifespan variability observed in the F2 experiment, an association should be found when looking

both at *mth*-like 11 and *mth*-like 5. For the latter gene, after the digestion of amplification product with *Bcl*I, two genotypes were found in this F2 association experiment: homozygous 0 (no digestion) (57.0±19.2 days; N=227), and heterozygous 1/0 (half digested) (60.2±22.1 days; N=200). It is not clear why no males homozygous for allele 1 were found, and this is in agreement with the results obtained in the previous association study, since in the AA cross was only observed one individual homozygous 1. However, and as previously reported in the first F2 association study, no association was found between this genotype and lifespan (Non-parametric Mann-Whitney Test; P>0.05).

The *GJ12490* ortholog, when digesting the amplification product with *Sau*3AI, we found three genotypes in this F2 experiment study: homozygous 0 (no digestion) (50.2±20.1 days; N=75), 1 (fully digested) (60.2±19.6 days; N=109) and heterozygous for alleles 0 and 1 (half-digested) (61.2±20.8 days; N=237). A significant association between this genotype and lifespan was found (Non-parametric Kruskal-Wallis Test; P<0.001). There are no significant differences regarding lifespan when genotypes 1 and 1/0 are compared (Non-parametric Mann-Whitney Test; P>0.05) but significant changes are observed when comparing genotypes 0 and 1/0 (Non-parametric Mann-Whitney Test, P<0.001), and when comparing genotypes 0 and 1 (Non-parametric Mann-Whitney Test; P<0.005). These results are significant after applying the sequential Bonferroni correction for multiple testing (P<0.05). Therefore, genotype 0 seems to be associated with short lifespan. Individuals with genotypes 1/1 and 1/0 show a 21.3% average increase in lifespan in comparison with individuals with genotype 0/0.

Nevertheless, as said before, more than 40 genes are described as being implicated in lifespan determination in *D. melanogaster*, so it is conceivable that variation at another gene in the neighborhood of *GJ12490* ortholog is the one showing variation that influences lifespan, and that this *mth*-like gene plays no role in the setting of lifespan.

3.5. The *mth*-like genes expression in *D. americana*

Once we have found out which *mth*-like genes might be involved in lifespan determination, the interest on their expression levels arose. In order to test if there are differences in the expression levels of each one of the seven genes analyzed, a qRT-PCR study was performed. The use of these seven genes tested by qRT-PCR was considered, once they have showed strong association with lifespan in at least one F2 association cross. The seven genes tested were *mth*-like 1, *GJ16328* ortholog, *mth*-like 11, *mth*-like 8, *mth*-like 9, *mth*-like 10 and *GJ12490* ortholog.

Since the aim of this experiment was to test the relationship between *mth*-like genes expression and lifespan, we used eight different sets of samples in order to measure the expression levels of these genes along time. To perform this study, only two strains were used – H5 and W11 – once they were the founder members of the cross where we saw the largest number of association between genotype and lifespan (AA cross). The expression levels were measured in males only, since the association studies were performed with males only. From the eight sets, four are for strain H5 and the other four are for W11 strain. The four kinds of sets for each strain correspond to four lifespan time points, i.e., 0 days, 10 days, 30 days and 60 days of life. With this sample it was possible to get a notion on how the expression of these genes varies over the life of the individuals from *D. americana* species.

RpL32 was used as an internal control. Nevertheless, the expression of this gene may vary with time and thus it is difficult to infer how the expression of *mth*-like genes varies along time. It should be noted that the concentration of RNA was normalized to 1µg upon the cDNA synthesis and the same cDNA was used to test the expression of all *mth*-like genes.

The results of qRT-PCR were analyzed with the use of $2^{-\Delta\Delta_{CT}}$ (Livak) method (Livak and Schmittgen, 2001). We only believe that the expression of these genes is different between the two strains (H5 and W11) when there are great differences (more than two-fold) that are consistent along the four lifespan time points.

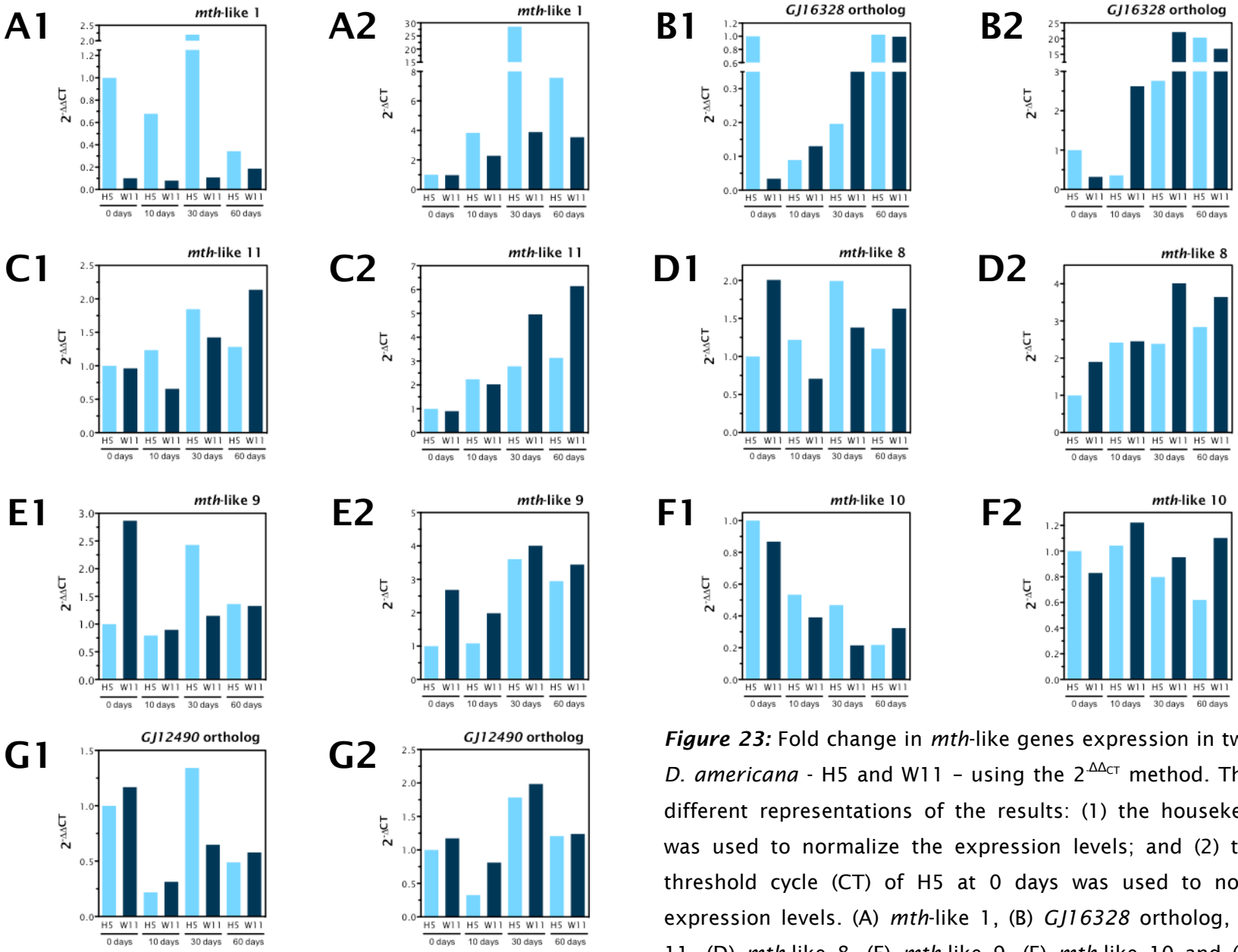

Figure 23: Fold change in *mth*-like genes expression in two strains of *D. americana* - H5 and W11 – using the $2^{-\Delta\Delta_{CT}}$ method. There are two different representations of the results: (1) the housekeeping gene was used to normalize the expression levels; and (2) the value of threshold cycle (CT) of H5 at 0 days was used to normalize the expression levels. (A) *mth*-like 1, (B) *GJ16328* ortholog, (C) *mth*-like 11, (D) *mth*-like 8, (E) *mth*-like 9, (F) *mth*-like 10 and (G) *GJ12490* ortholog.

The expression levels of *mth*-like 1 are higher in all times tested in H5 than W11 (figure 23 A). H5 individuals reach the highest level of expression of this *mth*-like gene at 30 days of life almost an 25-fold higher level than W11 at the same lifetime (figure 23 A2).

The ortholog of *GJ16328* presents a slight different behavior of expression (figure 23 B). The expression levels of this gene are similar between the two strains at 0 and 10 days, but at 30 days there is a difference of about 10-fold higher level in W11 than in H5 (figure 23 B2). However, at 60 days the expression level of this gene remains similar between the two strains.

When analyzing the expression levels of the other five genes tested – *mth*-like 11; *mth*-like 8; *mth*-like 9; *mth*-like 10 and ortholog of *GJ12490* -, we can see that there are no differences in their expression levels between the two strains in the four lifespan time points tested (figure 23 C-G). However, there are some evidences that should be retained. From all *D. americana mth*-like genes tested, the *mth*-like 10 has the lowest levels of expression throughout the life of individuals in both strains. On the other hand *mth*-like 1 and ortholog of *GJ16328* have the highest levels of expression.

Even though, it needs to be taken into account that these are only preliminary experiments, and that samples tested comprise the RNA obtained from three male individuals, in which could be included by chance an aberrant individual that carries a non-frequent allele.

4. DISCUSSION AND CONCLUSIONS

4.1. Dissecting *mth*-like genes family

The first aim of this study was trying to know how *mth*-like family of genes evolved through the years in Drosophila genus. We found out that Mth ectodomain, typical of Mth-like proteins, is only about 350-420 million years old and that it is likely to have arisen only once. And the way it is defined, this type of protein domain likely arose early in the Insect lineage. In Crustaceans, which are members of the sister group of insects, a protein sequence with a domain similar to *Drosophila* Mth-like proteins was found, but it is not recognizable as a Mth ectodomain. This protein proved to be hard to classify and might represent a proto-Mth protein.

Proteins with Mth typical features, including a recognizable Mth ectodomain with 10 typical cysteine residues and a 7tm domain (West *et al*., 2001), can be found in three Insect orders: Diptera, Leidoptera and Hemiptera. In addition to extended lifespan, *mth* mutants show higher wing-beat frequency and coordinated visuomotor entrainment to motion during simulated flight (Petrosyan *et al*., 2007). Taking into account that *mth*-like genes evolved early in the insect lineage and that *mth* mutants have a better performance during the flight, it is possible to infer that there is a possibility that *mth*-like genes are involved in the appearance and control of flight, the typical characteristic of Pterygota subclass that comprises Diptera, Lepidoptera and Hemiptera orders. Yet, we cannot put aside other scenarios that are equally probable. For example, wing flexion that is a typical feature of Neoptera is commonly described as a key innovation feature that allows insects to live in architecturally complex habitats (Mayhew, 2002). However, it is needed a better sampling from Insect groups that do not belong to Pterygota group in order to settle down this issue.

In *D. melanogaster,* sixteen *mth*-like genes were identified, one more than those identified by Nordström *et al*. (2009). Such as a large number of *mth*-like genes is unusual in insects and even within the *Drosophila* genus. Species from *melanogaster* subgroup show about twice as more *mth*-like genes than the other *Drosophila* species. Indeed *mth, mth*-like *2, mth*-like *3, mth*-like *4, mth*-like *6, mth*-like *7, mth*-like *12* and *mth*-like *13* are only present in species of the

melanogaster subgroup and this subgroup is about 10 million years old. Therefore, 50% of the *mth*-like genes found in *D. melanogaster* could be the resuld of a recent expansion of the *mth*-like gene family. Whether such an increase in gene number in species of *melanogaster* subgroup was driven by selection remains unknown. However, it should be noted that receptors, such as Mth, are expected to be among the most rarely retained gene duplications (Soyer and Creevey, 2010). Such high number of recent duplications may be possible if *mth*-like receptors are promiscuous, as suggested by Ja *et al*. (2009).

Whether the *mth*-like gene duplications observed in *melanogaster* subgroup was driven by positive selection should be considered in future work since at least both *mth* and *mth*-like 2 exhibit a significant pattern of adaptive amino acid divergence among *D. melanogaster*, *D. simulans* and *D. yakuba* (Duvernell *et al*., 2003).

We next address, whether the influence on lifespan determination is a feature of *mth* or whether it is a feature of the entire *mth*-like family. For that reason, we try to answer some questions like if the number of *mth*-like genes is the same in a species distantly related to *D. melanogaster,* once all the studies done till now that involve these genes were only performed in *D. melanogaster.* The model organism used is *D. americana*, a species that is diverging from *D. melanogaster* for about 40 million years (Russo *et al*., 1995). We looked for the *mth*-like genes that are present in its genome. As said in the results chapter, we found nine *mth*-like genes in this species, which is closely related to *D. virilis*. The *mth*-like family of *D. americana* has a lower number of members than *D. melanogaster.* Moreover, both *D. virilis* and *D. americana* do not have an ortholog gene of *mth*. However, these species have *mth*-like genes that have the Mth typical features – Mth ectodomain with ten conserved cysteine residues and a 7tm domain (West *et al*., 2001).

Nevertheless, this high diversity of *mth*-like genes with different characteristics make this study more reliable, since the aim of this experiment was to find out if all *mth*-like genes are involved in lifespan determination of a species distantly related to *D. melanogaster.* Moreover, this is also important, once few is known about *mth*-like genes and one thing known is that Sun peptides that are able to activate Mth receptor cannot activate other Mth-like proteins (Cvejic *et al*., 2004).

4.2. The association between *mth*-like genes and lifespan

Naturally occurring amino acid variation at the *D. melanogaster mth* gene was previously associated with lifespan differences (Schmidt *et al.*, 2000, Duvernell *et al.*, 2003). For that reason we searched for polymorphisms that could explain why there are flies that live only 30 days and others that live almost 120 days. The protein encoded by *mth* gene seems to have many functions or at least seems to have an important role in many pathways that are important for the life of an individual. Most of the proposed functions of this gene involve its participation in nervous system of the fly, such as control of synaptic efficiency at glutamatergic neuromuscular junctions (Song *et al.*, 2002), regulation of stress resistance (Lin *et al.*, 1998) and enhancement of some sensorial abilities (Petrosyan *et al.*, 2007). These processes are very important for the fly in order to be aware of everything that is happening around them, like predators, food, weather and so one. Moreover, if the synaptic efficiency is not well controlled it can lead to a decrease or excessive transmissition at synapses that will compromise the fly's life. Therefore, Mth-like proteins could be involved in lifespan determination.

The first association study performed gave us some important information about the *mth*-like family members and their association with lifespan. These results support the view that *mth*-like genes with Mth ectodomain are likely to be involved in lifespan determination. The five crosses comprised in this association study allow us to test the importance of *mth*-like genes in lifespan determination of a sample that include individuals originating from strains that represent the distribution of *D. americana* species (Patterson and Stone, 1952). From the nine *mth*-like genes of this species, only three of them do not show association with lifespan in AA cross, and two of these three genes do not have the Mth typical features. These results suggest that in order to influence lifespan; the Mth-like protein should have an Mth ectodomain.

However, in AA cross, a *mth*-like gene without Mth ectodomain showed a strong association with lifespan. This gene, *mth*-like 1, has a long 7tm domain and is localized in chromosome *X*. It could be that this gene is not well annotated which then will lead to a wrong prediction of the conserved domains present in the protein. In the future, this gene should be tested in females too in order to understand if this association with lifespan remains strong when there are three possible genotypes.

In relation to the two *mth*-like genes that did not show association with lifespan in any of the crosses tested – *mth*-like 5 and *mth*-like 14 –, they have a very simple protein structure without any conserved domain. These two *mth*-like genes together with *mth*-like 1 are the eldest members of *mth*-like family.

Besides these two genes, there is a *mth*-like gene that has all the typical features of Mth protein and did not show association with lifespan determination in AA cross but showed a strong association in NA cross – *mth*-like 10. In the future, there is a need of looking for the expression levels of this gene in O57 strain, since together with H5 strain they founded NA cross.

Since the beginning of this study, the possibility that the influence of this family of genes in lifespan determination could be due to differences in the expression levels of *mth*-like genes has not been discarded. For that reason, we tested the expression level of the genes that showed association with lifespan at four lifespan time points. Despite *mth*-like 1, none of the other six genes tested showed consistent differences in the expression levels between strains H5 and W11. These results suggest that the influence of these genes in lifespan might be not due to differences in their expression levels, but to amino acid polymorphisms that could influence the function of the protein.

However, as the expression levels were only tested in males, in future work, it will be important to test the expression levels in a different way. It will be important to extract RNA of ten males and ten females from about four different strains of *D. americana* individually in about fifteen lifespan time points – 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 and 70 days. This will allow performing a more accurate study of the expression level of these genes throughout the life of *D. americana* individuals and will clarify if the expression levels of *mth*-like 1 will also be different between strains in females. Moreover, the expression levels of these genes seem not be determinant for lifespan determination, but they might be important in some points of fly's life, like when it is under some stress (Lin *et al.*, 1998, Petrosyan *et al.*, 2007). For that reason, it might be important to put some individuals under some kind of stress like presence of paraquat, high temperature or confining the fly to a small tube where it can barely beat its wings during a limited time and test the expression levels of these genes after the exposure to the stress. These experiments will reveal whether these genes are important for lifespan determination and if they are important in some points of

the fly's life that allow them to overtake some difficulties and consequently live longer.

Another approach that could be made in order to understand the real association between *mth*-like genes and lifespan is to study the influence of amino acid polymorphism on lifespan, using wild-caught individuals, although it requires a large number of individuals.

Besides these experiments, it is also important to search for the function that these genes play. Although determining the crystal structure of a transmembrane protein is a really difficult job, in the future we can use molecular modeling in order to study the active site of protein Mth and try to understand what kind of ligand is able to bind inside it. With this type of study, we will also be able to design Mth inhibitors and consequently see what the inhibition of this receptor causes in the individuals.

With this work, we can conclude that *mth*-like gene family does not have a constant number of members inside *Drosophila* genus. However, *mth*-like genes that have Mth ectodomain showed a strong association with lifespan in *D. americana*, a distantly related species with *D. melanogaster.* From these results, we can suggest that *mth*-like genes with Mth ectodomain might be associated with lifespan determination in all orders of insects where they have been identified: Diptera, Lepidoptera and Hemiptera. *mth*-like genes thus seem to be true candidate genes that might explain a large part of lifespan.

5. REFERENCES

ADAMS, M., CELNIKER, S., HOLT, R., EVANS, C., GOCAYNE, J., AMANATIDES, P., SCHERER, S., LI, P., HOSKINS, R., GALLE, R., GEORGE, R., LEWIS, S., RICHARDS, S., ASHBURNER, M., HENDERSON, S., SUTTON, G., WORTMAN, J., YANDELL, M., ZHANG, Q., CHEN, L., BRANDON, R., ROGERS, Y., BLAZEJ, R., CHAMPE, M., PFEIFFER, B., WAN, K., DOYLE, C., BAXTER, E., HELT, G., NELSON, C., GABOR, G., ABRIL, J., AGBAYANI, A., AN, H., ANDREWS-PFANNKOCH, C., BALDWIN, D., BALLEW, R., BASU, A., BAXENDALE, J., BAYRAKTAROGLU, L., BEASLEY, E., BEESON, K., BENOS, P., BERMAN, B., BHANDARI, D., BOLSHAKOV, S., BORKOVA, D., BOTCHAN, M., BOUCK, J., BROKSTEIN, P., BROTTIER, P., BURTIS, K., BUSAM, D., BUTLER, H., CADIEU, E., CENTER, A., CHANDRA, I., CHERRY, J., CAWLEY, S., DAHLKE, C., DAVENPORT, L., DAVIES, P., DE PABLOS, B., DELCHER, A., DENG, Z., MAYS, A., DEW, I., DIETZ, S., DODSON, K., DOUP, L., DOWNES, M., DUGAN-ROCHA, S., DUNKOV, B., DUNN, P., DURBIN, K., EVANGELISTA, C., FERRAZ, C., FERRIERA, S., FLEISCHMANN, W., FOSLER, C., GABRIELIAN, A., GARG, N., GELBART, W., GLASSER, K., GLODEK, A., GONG, F., GORRELL, J., GU, Z., GUAN, P., HARRIS, M., HARRIS, N., HARVEY, D., HEIMAN, T., HERNANDEZ, J., HOUCK, J., HOSTIN, D., HOUSTON, K., HOWLAND, T., WEI, M., IBEGWAM, C., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science,* 287**,** 2185-2280.

ASHBURNER, M. & BERGMAN, C. 2005. *Drosophila melanogaster*: a case study of a model genomic sequence and its consequences. *Genome Research,* 15**,** 1661-1668.

BRODY, T. & CRAVCHIK, A. 2000. *Drosophila melanogaster* G protein-coupled receptors. *The Journal of Cell Biology,* 150**,** F83-8.

CVEJIC, S., ZHU, Z., FELICE, S. J., BERMAN, Y. & HUANG, X. Y. 2004. The endogenous ligand Stunted of the GPCR Methuselah extends lifespan in *Drosophila*. *Nature Cell Biology,* 6**,** 540-6.

DORSAM, R. & GUTKIND, J. 2007. G-protein-coupled receptors and cancer. *Nature Reviews. Cancer,* 7**,** 79-173.

DUVERNELL, D. D., SCHMIDT, P. S. & EANES, W. F. 2003. Clines and adaptive evolution in the *methuselah* gene region in *Drosophila melanogaster*. *Molecular Ecology,* 12**,** 1277-85.

FAN, Y., SUN, P., WANG, Y., HE, X., DENG, X., CHEN, X., ZHANG, G., CHEN, X. & ZHOU, N. 2010. The G protein-coupled receptors in the silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology,* 40**,** 581-91.

FREDRIKSSON, R., LAGERSTROM, M. C., LUNDIN, L. G. & SCHIOTH, H. B. 2003. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Molecular Pharmacology,* 63**,** 1256-72.

FREDRIKSSON, R. & SCHIÖTH, H. 2005. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Molecular Pharmacology,* 67**,** 1414-1439.

GAUNT, M. & MILES, M. 2002. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Molecular Biology and Evolution,* 19**,** 748-809.

GIANNAKOU, M. E. & PARTRIDGE, L. 2007. Role of insulin-like signalling in *Drosophila* lifespan. *Trends in Biochemical Sciences,* 32**,** 180-8.

HILL, C. A., FOX, A. N., PITTS, R. J., KENT, L. B., TAN, P. L., CHRYSTAL, M. A., CRAVCHIK, A., COLLINS, F. H., ROBERTSON, H. M. & ZWIEBEL, L. J. 2002. G protein-coupled receptors in *Anopheles gambiae*. *Science,* 298**,** 176-8.

HILTON, H. & HEY, J. 1996. DNA sequence variation at the period locus reveals the history of species and speciation events in the *Drosophila virilis* group. *Genetics,* 144**,** 1015-25.

HILTON, H. & HEY, J. 1997. A multilocus view of speciation in the *Drosophila virilis* species group reveals complex histories and taxonomic conflicts. *Genetical Research,* 70**,** 185-379.

HOLZ, G. & HABENER, J. 1998. Black widow spider alpha-latrotoxin: a presynaptic neurotoxin that shares structural homology with the glucagon-like peptide-

1 family of insulin secretagogic hormones. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology* 121**,** 177-261.

HORN, F., WEARE, J., BEUKERS, M. W., HORSCH, S., BAIROCH, A., CHEN, W., EDVARDSEN, O., CAMPAGNE, F. & VRIEND, G. 1998. GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Research,* 26**,** 275-9.

HSU, T. C. 1952. Chromosomal variation and evolution in the *virilis* group of *Drosophila*. *University of Texas Publications*, 5204, 35-72.

HUGHES, R. D. 1939. An Analysis of the chromosomes of the two sub-species *Drosophila virilis virilis* and *Drosophila virilis americana*. *Genetics,* 24**,** 811-34.

JA, W. W., CARVALHO, G. B., MADRIGAL, M., ROBERTS, R. W. & BENZER, S. 2009. The Drosophila G protein-coupled receptor, Methuselah, exhibits a promiscuous response to peptides. *Protein Science,* 18**,** 2203-8.

JAAKOLA, V.-P., GRIFFITH, M., HANSON, M., CHEREZOV, V., CHIEN, E., LANE, J., IJZERMAN, A. & STEVENS, R. 2008. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science (New York, N.Y.),* 322**,** 1211-1218.

JOHNSON, E., BOHN, L. & TAGHERT, P. 2004. *Drosophila CG8422* encodes a functional diuretic hormone receptor. *The Journal of Experimental Biology,* 207**,** 743-751.

KRASNOPEROV, V., BITTNER, M., BEAVIS, R., KUANG, Y., SALNIKOW, K., CHEPURNY, O., LITTLE, A., PLOTNIKOV, A., WU, D., HOLZ, R. & PETRENKO, A. 1997. alpha-Latrotoxin stimulates exocytosis by the interaction with a neuronal G-protein-coupled receptor. *Neuron,* 18**,** 925-962.

KRUPNICK, J. G., GUREVICH, V. V. & BENOVIC, J. L. 1997. Mechanism of quenching of phototransduction. Binding competition between arrestin and transducin for phosphorhodopsin. *The Journal of Biological Chemistry,* 272**,** 18125-31.

KUMAR, S., NEI, M., DUDLEY, J. & TAMURA, K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics,* 9**,** 299-605.

LAGERSTROM, M. C. & SCHIOTH, H. B. 2008. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nature Reviews. Drug Discovery,* 7**,** 339-57.

LIN, Y. J., SEROUDE, L. & BENZER, S. 1998. Extended life-span and stress resistance in the *Drosophila* mutant *methuselah*. *Science,* 282**,** 943-6.

LIVAK, K. & SCHMITTGEN, T. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-410.

MARCHLER-BAUER, A., LU, S., ANDERSON, J., CHITSAZ, F., DERBYSHIRE, M., DEWEESE-SCOTT, C., FONG, J., GEER, L., GEER, R., GONZALES, N., GWADZ, M., HURWITZ, D., JACKSON, J., KE, Z., LANCZYCKI, C., LU, F., MARCHLER, G., MULLOKANDOV, M., OMELCHENKO, M., ROBERTSON, C., SONG, J., THANKI, N., YAMASHITA, R., ZHANG, D., ZHANG, N., ZHENG, C. & BRYANT, S. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research,* 39**,** 9.

MATTHEWS, K., KAUFMAN, T. & GELBART, W. 2005. Research resources for *Drosophila*: the expanding universe. *Nature Reviews. Genetics,* 6**,** 179-272.

MAYHEW, P. J. 2002. Shifts in hexapod diversification and what Haldane could have said. *Proceedings. Biological Sciences / The Royal Society* 269**,** 969-74.

MCALLISTER, B. F. 2002. Chromosomal and allelic variation in *Drosophila americana:* selective maintenance of a chromosomal cline. *Genome,* 45**,** 13-21.

MCALLISTER, B. F. & CHARLESWORTH, B. 1999. Reduced sequence variability on the Neo-Y chromosome of *Drosophila americana americana*. *Genetics,* 153**,** 221-33.

MCALLISTER, B. F. & MCVEAN, G. A. 2000. Neutral evolution of the sex-determining gene transformer in *Drosophila*. *Genetics,* 154**,** 1711-20.

MCALLISTER, B. F., SHEELEY, S. L., MENA, P. A., EVANS, A. L. & SCHLOTTERER, C. 2008. Clinal distribution of a chromosomal rearrangement: a precursor to chromosomal speciation? *Evolution,* 62**,** 1852-65.

MCCUDDEN, C. R., HAINS, M. D., KIMPLE, R. J., SIDEROVSKI, D. P. & WILLARD, F. S. 2005. G-protein signaling: back to the future. *Cellular and Molecular Life Sciences,* 62**,** 551-77.

MEUSEMANN, K., VON REUMONT, B., SIMON, S., ROEDING, F., STRAUSS, S., KÜCK, P., EBERSBERGER, I., WALZL, M., PASS, G., BREUERS, S., ACHTER, V., VON HAESELER, A., BURMESTER, T., HADRYS, H., WÄGELE, J. & MISOF, B. 2010. A phylogenomic approach to resolve the arthropod tree of life. *Molecular Biology and Evolution,* 27**,** 2451-2515.

MIGEOTTE, I., COMMUNI, D. & PARMENTIER, M. 2006. Formyl peptide receptors: a promiscuous subfamily of G protein-coupled receptors controlling immune responses. *Cytokine Growth Factor Rev,* 17**,** 501-520.

MOCKETT, R. J. & SOHAL, R. S. 2006. Temperature-dependent trade-offs between longevity and fertility in the *Drosophila* mutant, *methuselah*. *Experimental Gerontology,* 41**,** 566-73.

MORALES-HOJAS, R., REIS, M., VIEIRA, C. P. & VIEIRA, J. 2011. Resolving the phylogenetic relationships and evolutionary history of the *Drosophila virilis* group using multilocus data. *Molecular Phylogenetics and Evolution,* 60**,** 249-58.

MORALES-HOJAS, R., VIEIRA, C. P. & VIEIRA, J. 2008. Inferring the evolutionary history of *Drosophila americana* and *Drosophila novamexicana* using a multilocus approach and the influence of chromosomal rearrangements in single gene analyses. *Molecular Ecology,* 17**,** 2910-26.

NEUBIG, R. R. 2002. Regulators of G protein signaling (RGS proteins): novel central nervous system drug targets. *The Journal of Peptide Research,* 60**,** 312-6.

NORDSTRÖM, K. J., LAGERSTROM, M. C., WALLER, L. M., FREDRIKSSON, R. & SCHIOTH, H. B. 2009. The Secretin GPCRs descended from the family of Adhesion GPCRs. *Molecular Biology and Evolution,* 26**,** 71-84.

PAABY, A. B. & SCHMIDT, P. S. 2008. Functional significance of allelic variation at *methuselah*, an aging gene in *Drosophila*. *PLoS One,* 3**,** e1987.

PAABY, A. B. & SCHMIDT, P. S. 2009. Dissecting the genetics of longevity in *Drosophila melanogaster*. *Fly,* 3**,** 29-38.

PARKES, T., ELIA, A., DICKINSON, D., HILLIKER, A., PHILLIPS, J. & BOULIANNE, G. 1998. Extension of *Drosophila* lifespan by overexpression of human SOD1 in motorneurons. *Nature Genetics,* 19**,** 171-175.

PATTERSON, J. T. & STONE, W. S. 1952. Evolution in the genus *Drosophila*. *Macmillan New York*.

PETROSYAN, A., HSIEH, I. H. & SABERI, K. 2007. Age-dependent stability of sensorimotor functions in the life-extended *Drosophila* mutant *methuselah*. *Behavior Genetics,* 37**,** 585-94.

PIERCE, K., PREMONT, R. & LEFKOWITZ, R. 2002. Seven-transmembrane receptors. *Nature Reviews. Molecular Cell Biology,* 3**,** 639-689.

PROMISLOW, D. E., TATAR, M., KHAZAELI, A. A. & CURTSINGER, J. W. 1996. Age-specific patterns of genetic variance in *Drosophila melanogaster*. I. Mortality. *Genetics,* 143**,** 839-48.

REIS, M., VIEIRA, C. P., MORALES-HOJAS, R., AGUIAR, B., ROCHA, H., SCHLOTTERER, C. & VIEIRA, J. 2011. A comparative study of the short term cold resistance response in distantly related *Drosophila* species: the role of *regucalcin* and *frost*. *PLoS One,* 6**,** e25520.

REIS, M., VIEIRA, C. P., MORALES-HOJAS, R. & VIEIRA, J. 2008. An old *bilbo*-like non-LTR retroelement insertion provides insight into the relationship of species of the *virilis* group. *Gene,* 425**,** 48-55.

RICHARDS, S., LIU, Y., BETTENCOURT, B., HRADECKY, P., LETOVSKY, S., NIELSEN, R., THORNTON, K., HUBISZ, M., CHEN, R., MEISEL, R., COURONNE, O., HUA, S., SMITH, M., ZHANG, P., LIU, J., BUSSEMAKER, H., VAN BATENBURG, M.,

HOWELLS, S., SCHERER, S., SODERGREN, E., MATTHEWS, B., CROSBY, M., SCHROEDER, A., ORTIZ-BARRIENTOS, D., RIVES, C., METZKER, M., MUZNY, D., SCOTT, G., STEFFEN, D., WHEELER, D., WORLEY, K., HAVLAK, P., DURBIN, K., EGAN, A., GILL, R., HUME, J., MORGAN, M., MINER, G., HAMILTON, C., HUANG, Y., WALDRON, L., VERDUZCO, D., CLERC-BLANKENBURG, K., DUBCHAK, I., NOOR, M., ANDERSON, W., WHITE, K., CLARK, A., SCHAEFFER, S., GELBART, W., WEINSTOCK, G. & GIBBS, R. 2005. Comparative genome sequencing of *Drosophila pseudoobscura:* chromosomal, gene, and cis-element evolution. *Genome Research,* 15**,** 1-19.

RONQUIST, F. & HUELSENBECK, J. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-1576.

ROSENBAUM, D., RASMUSSEN, S. & KOBILKA, B. 2009. The structure and function of G-protein-coupled receptors. *Nature,* 459**,** 356-419.

RUSSO, C. A., TAKEZAKI, N. & NEI, M. 1995. Molecular phylogeny and divergence times of Drosophilid species. *Molecular Biology and Evolution,* 12**,** 391-404.

SCHAFER, M. A., ORSINI, L., MCALLISTER, B. F. & SCHLOTTERER, C. 2006. Patterns of microsatellite variation through a transition zone of a chromosomal cline in *Drosophila americana*. *Heredity,* 97**,** 291-5.

SCHIÖTH, H., NORDSTRÖM, K. & FREDRIKSSON, R. 2007. Mining the gene repertoire and ESTs for G protein-coupled receptors with evolutionary perspective. *Acta physiologica,* 190**,** 21-52.

SCHMIDT, P. S., DUVERNELL, D. D. & EANES, W. F. 2000. Adaptive evolution of a candidate gene for aging in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 10861-5.

SCHRODINGER, LLC 2010. The PyMOL Molecular Graphics System, Version 1.3r1.

SONG, W., RANJAN, R., DAWSON-SCULLY, K., BRONK, P., MARIN, L., SEROUDE, L., LIN, Y. J., NIE, Z., ATWOOD, H. L., BENZER, S. & ZINSMAIER, K. E. 2002. Presynaptic regulation of neurotransmission in *Drosophila* by the G protein-coupled receptor *methuselah*. *Neuron,* 36**,** 105-19.

SOYER, O. S. & CREEVEY, C. J. 2010. Duplicate retention in signalling proteins and constraints from network dynamics. *Journal of Evolutionary Biology,* 23**,** 2410-21.

SPICER, G. S. 1992. Reevaluation of the phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae). *Annals of the Entomological Society of America,* 85**,** 11-36.

SPICER, G. S. & BELL, C. D. 2002. Molecular phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae) inferred from mitochondrial 12S and 16S ribosomal RNA genes. *Annals of the Entomological Society of America,* 95**,** 156-317.

STRADER, C. D., FONG, T. M., TOTA, M. R., UNDERWOOD, D. & DIXON, R. A. 1994. Structure and function of G protein-coupled receptors. *Annual Review of Biochemistry,* 63**,** 101-32.

STROTMANN, R., SCHRÖCK, K., BÖSELT, I., STÄUBERT, C., RUSS, A. & SCHÖNEBERG, T. 2011. Evolution of GPCR: change and continuity. *Molecular and Cellular Endocrinology,* 331**,** 170-178.

THROCKMORTON, L. 1982. The *virilis* species group. *The Genetics and Biology of Drosophila,* 3**,** 227-523.

VIEIRA, C., ALMEIDA, A., DIAS, J. & VIEIRA, J. 2006. On the location of the gene(s) harbouring the advantageous variant that maintains the *X/4* fusion of *Drosophila americana*. *Genetical Research,* 87**,** 163-237.

VIEIRA, C., COELHO, P. & VIEIRA, J. 2003. Inferences on the evolutionary history of the *Drosophila americana* polymorphic *X/4* fusion from patterns of polymorphism at the X-linked *paralytic* and *elav* genes. *Genetics,* 164**,** 1459-1528.

VIEIRA, J., MCALLISTER, B. F. & CHARLESWORTH, B. 2001. Evidence for selection at the *fused1* locus of *Drosophila americana*. *Genetics,* 158**,** 279-90.

WALLENFANG, M. R., NAYAK, R. & DINARDO, S. 2006. Dynamics of the male germline stem cell population during aging of *Drosophila melanogaster*. *Aging Cell,* 5**,** 297-304.

WELLENDORPH, P., HANSEN, K., BALSGAARD, A., GREENWOOD, J., EGEBJERG, J. & BRÄUNER-OSBORNE, H. 2005. Deorphanization of GPRC6A: a promiscuous L-alpha-amino acid receptor with preference for basic amino acids. *Molecular Pharmacology,* 67**,** 589-686.

WEST, A. P., JR., LLAMAS, L. L., SNOW, P. M., BENZER, S. & BJORKMAN, P. J. 2001. Crystal structure of the ectodomain of Methuselah, a *Drosophila* G protein-coupled receptor associated with extended lifespan. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 3744-9.

WHEELER, M. 1981. The Drosophilidae: a taxonomic overview. *The Genetics and Biology of Drosophila,* 3**,** 1-98.

WHITTAKER, C. A., BERGERON, K. F., WHITTLE, J., BRANDHORST, B. P., BURKE, R. D. & HYNES, R. O. 2006. The echinoderm adhesome. *Developmental Biology,* 300**,** 252-66.

WILDEN, U. 1995. Duration and amplitude of the light-induced cGMP hydrolysis in vertebrate photoreceptors are regulated by multiple phosphorylation of rhodopsin and by arrestin binding. *Biochemistry,* 34**,** 1446-54.

XIA, X. & XIE, Z. 2001. DAMBE: software package for data analysis in molecular biology and evolution. *The Journal of Heredity,* 92**,** 371-374.